

ASHLEY RINDSBERG

How AI Is Being Weaponized Against Jews

With Wikipedia as its muse



Y JANUARY 2025, a group of anti-Israel editors on Wikipedia had pulled off a feat that even Kremlin propagandists could only dream of. Operating with striking discipline, the “Gang of 40,” as I dubbed them, zeroed in on the most valuable targets in the information war of the Israeli–Palestinian conflict, above all the Wikipedia entry on Zionism.

The Gang of 40 set out to equate Zionism with ethnic cleansing. They focused their energy on the article’s lead section, from which Google pulls information that populates its search results. Even more critically, it’s the section that AI models largely draw their reference data from.

To say the operation was a success would wildly understate it. Prior to October 7, Wikipedia defined Zionism as “a nationalist movement that emerged in the 19th century to espouse support for the

establishment of a homeland for the Jewish people in Palestine.” But now, after the Gang of 40 spent hours warping the article, it reads, “Zionism is an ethnocultural nationalist movement that emerged in late 19th-century Europe to establish and support a Jewish homeland through colonization in the region of Palestine.” The second sentence, added by a single editor and cemented through weeks of endless browbeating and sheer doggedness, struck an even greater blow: “Zionists wanted to create a Jewish state in Palestine with as much land, as many Jews, and as few Palestinian Arabs as possible.”

This alone—reducing Zionism to the false claim that it was designed to cleanse the land of Arabs—would have constituted a triumph. But the coup de grâce came later. Commandeering Wikipedia’s conceptual machinery, the group instituted a “moratorium” on the lead, locking it in place. (In Wikipedia’s governance framework, a moratorium is an informal, ill-defined practice. In this case, the concept was applied to the entire lead of an article—something that veteran editors I spoke with do not recall ever seeing before.)

The effect was that, for a full year, no one would be able to touch even a single word of the *Zionism* entry lead. But the duration of the moratorium was just as important as its scope. In the winter of 2025, the top echelon of the Gang of 40 leadership would be banned by Wikipedia from editing in the Palestine-Israel articles topic space. The likely length of their ban? Twelve months—that is, only for as long as everyone else was locked out.

The episode was a dramatic escalation in the ongoing, ever-intensifying narrative war playing out online over Israel and the Jewish people. It is by no means unique. The Wikipedia entry entitled “Gaza genocide,” captured by the same group, now frames an alleged and highly contested notion of genocide as settled fact. They have deleted terror attacks committed by Hezbollah,

scrubbed mentions of Hamas's 1988 genocidal charter from dozens of articles, made anachronistic alterations — such as changing the title of Wikipedia's "Hellenistic Judea" entry to "Hellenistic Palestine" — and removed photos of the Dead Sea Scrolls from the entry on Israel.

Dig into Gang of 40 activities, or activities of the new guard that has stepped in to continue their work of historical revisionism, and you will be astonished by the breadth of their activity. No issue, no topic, is too small or trivial to evade their ideological attention. Kibbutzim and moshavim you've never heard of, the entry on the "Jaffa orange" ("significant ethnic cleansing"), and "Jewish Insider" — all have been under the editorial scalpel of this surgical group.

By now, it's clear to most that a fierce information war is under way on platforms such as Facebook, Instagram, and X. Yet it's at a deeper layer of what we call the "knowledge infrastructure" where quiet, slow efforts have lasting impact. Claims embedded in platforms whose data disproportionately feed AI—notably Wikipedia and Reddit—get crystallized into what AI systems treat as verified fact or knowledge.

These seed platforms are ripe for manipulation in ways that Facebook, Instagram, and X are not. While social media companies have either forced or strongly incentivized user identity verification, there is no such process on Wikipedia. Editors shaping information about some of the most contentious and sensitive topics remain anonymous. The identities of the site's most powerful administrators, including those with the ability to search and identify the IP address of any user, remain unknown. While X and Facebook spend vast sums employing content moderators, Wikipedia employs none. Its own users are tasked with the job of policing truth, and needless to say, they don't all agree on what it is. As for Reddit, there too the task of content moderation is largely left to users.



Anonymity is a key feature of digital life. Many people lurk and troll online behind fake accounts. It's one thing to do this while posting comments on Facebook and Instagram. But doing so on a platform that has the power to determine public perceptions of what's true and false—and that feeds this data into the AI systems shaping our collective future—is a propagandist's dream.

Case in point: A Hamas-linked network operating on Reddit (which I exposed last year) launders messaging from Telegram channels run by Hamas, Hezbollah, the Houthis, and numerous other U.S.-designated foreign terror organizations onto every mainstream social media platform. Because of Reddit's agreements with companies such as OpenAI, the data end up in AI training corpuses. Reddit refuses to curb these activities, even though the company has taken an aggressive position on banning communities posting anti-trans or fat-shaming perspectives—and took the extraordinary step of banning the biggest pro-Trump community online only months before the 2020 election. Clearly they're not free-speech absolutists.

The core of the problem is that we, Jews and society as a whole, have not yet assimilated the idea that narrative threats have now become as dangerous as any cyber threat. The world of cybersecurity speaks about “zero-day” vulnerabilities—risks that are clear and present, with no time delay between a threat and its execution. But that is also the nature of the threats we're facing in the information pipeline feeding AI systems, with one big difference: The narrative threats are not merely potential but constantly executed. They are exploited every day, at scale, with stunning efficiency.

State-aligned propagandists have not overlooked this generational opportunity to reshape the information battleground. For years a small group of editors has worked to scrub Wikipedia articles of men-

tions of human rights abuses committed by the Iranian regime and its top leaders. Several weeks after the regime in Tehran slaughtered thousands of Iranians in the 2025–2026 protests, Wikipedia’s media repository, WikiCommons, was flooded with thousands of images and videos produced by Tasnim News Agency (a U.S.-sanctioned entity), Mehr News Agency (which is state-owned), and Khamenei.ir (the official website of the then-supreme leader). My research has revealed that Iranian state media have been cited more than 20,000 times on Wikipedia as reliable sources in hundreds of articles. Outlets linked to Hamas, Hezbollah, and Palestinian Islamic Jihad are cited more than 8,000 times. In multiple cases, content had been copied and pasted directly into Wikipedia entries from Islamist terror groups such as Palestinian Islamic Jihad. There were no fewer than 100 citations using al-Qaeda-linked outlets as sources. In one particularly egregious case, a Wikipedia administrator publicly directed a jihad-curious user to a Muslim Brotherhood recruitment site.

If Wikimedia Foundation (WMF), the NGO that owns and operates Wikipedia, cared to know about the kind of content flowing through its site, it would. WMF operates on a \$200 million annual budget, more than enough to implement monitoring and moderation technologies and to employ hundreds of product and engineering team members. It prefers not to know because knowing would make it accountable for this content. Like every other major social media platform, it has fought hard to resist that responsibility.

In the meantime, influence campaigns such as the Gang of 40 and Reddit’s r/Palestine group are running rampant. We’re starting to see the effects of this not just in retrieval data (the information that chatbots pull in response to a prompt or query) but, more worryingly, in what we might describe as the worldview of AI systems, which are being shaped by the training data in ways that are less visible but far more consequential.



The question many are asking is what should be done. But that question contains a prior failure. You can't answer "What should be done?" when you don't yet know what's happening.

Even after years of investigation, I suspect that I have, at best, 3 percent visibility into the scope of what's happening on knowledge infrastructure platforms. Three percent. That's not modesty—it's a measurement of how completely these operations have succeeded in making themselves invisible. The Gang of 40 operated for years before anyone named or noticed them. The Hamas-linked Reddit pipeline I exposed had been running, largely undetected, across every major platform since at least October 7. The Iranian state-media flood of Wikipedia happened in near-total silence. What I've documented is not the tip of the iceberg. It's a floating chip off that tip. Nearly three years after October 7, this is a terrifying and untenable state of affairs.

At a baseline, the Jewish world needs to internalize the fact that we are in the middle of a generational information war, in which the most established facts are turned into bald-faced lies. Joel Finkelstein, founder of the Network Contagion Research Institute and one of the most rigorous practitioners working on these issues, put the stakes plainly to me: "Every Jewish communal investment you make is downstream of whether we win the narrative war. If we lose that, the leadership programs and yeshivas have no safe environment to operate in."

He's right. And the corollary is just as stark: You cannot prevail on a battlefield you haven't even mapped. There is an urgent need for the basic investigative work of understanding what we're actually facing, who is doing it, how it works, at what scale, through which networks, and—most critically—with whose money.

That last point is key. We have strong signals—and, in some cases,

outright evidence—that the vast network of information operations working against Israel and the Jewish people is funded to a significant extent by state-level actors, including those affiliated with Russia, China, Iran, Turkey, and Qatar. This funding carries nearly no risk of running afoul of legal, compliance, or regulatory issues and is all but impervious to exposure. It's a weaponization of the very freedoms afforded by Western nations prohibited domestically by these same countries. They have captured the narrative not just about Jews and Israel but—far more worryingly—about the United States. They are using open knowledge platforms to poison AI systems, and they're doing so with impunity.

The most significant state-backed narrative operations hide in plain sight: Al Jazeera, RT, *China Daily*, Anadolu Agency, Tasnim. These outlets freely seed anti-American propaganda in the United States while their own governments clamp down on good-faith domestic and international journalism within their own borders. The formula, stunningly effective, is also maddeningly simple: lies abroad, repression at home.

The most important step is to undertake investigative initiatives capable of exposing state-backed and state-linked narrative networks. We see the same root cause behind the phenomenon of obscure influencers who shoot to prominence after uttering the crudest anti-Israel or anti-Jewish propaganda, with PR agencies pitching them to podcasts, and law firms rushing to defend them when they meet legal trouble.

The current strategy seems to be to react to every X post with a torrent of outrage (guilty as charged). But that's merely suicide by a thousand paper cuts. Each furious quote-tweet serves to amplify the poisonous messaging. What we need is robust journalistic, financial, and technical investigation capable of picking apart these networks to expose how Americans are being served tainted media every min-

ute of every day. We need to inform audiences that, far from organic, the info they're seeing is the product of foreign interference. We need social media and knowledge platforms to call this out, as X did when it started showing the location of accounts last November. And we need Congress to act. When lawmakers approved Section 230—the provision of the Telecommunications Act of 1996 that created a legal carve-out for internet service providers that has shielded online platforms from accountability for what appears on them—the internet was in its infancy. There was a real possibility that a flurry of lawsuits seeking to hold online platforms liable for anonymous posts by third parties could strangle the nascent web. More important, online information warfare was not even a gleam in the eye of the most forward-thinking CCP or Kremlin propagandist.

Three decades later, our information environments more closely resemble war zones than the digital idylls envisioned by the online pioneers. And far from a fragile, tenuous thing, the internet has become the primary setting of our commercial, social, legal, and educational lives. Free speech online must be preserved and protected, that much is evident; but one of the many great lessons we learn from the Founders of the American Republic is that freedom is possible only when we are first freed from the tyranny of anarchy.



Today, anti-American, anti-Western, anti-Israel, and anti-Jewish information networks operate with asymmetric advantages. They choose the targets, dictate the tempo, coordinate the narratives across platforms, and exploit institutional fragmentation on the other side. Over decades, they have constructed a narrative infrastructure capable of rallying millions by invoking single words—*genocide*, *apartheid*, *colonialism*—no matter how spurious or even ridiculous.

The Jewish world, on the other hand, faces all these risks and constraints—and many more. We are endlessly busy turning out our pockets to prove we're not thieves. We have no buzzword narratives. You can't explain in a tweet the deep and complex integration of Israeli Arabs into the country's broader society. You can't break down Hamas's decades-long strategy of ideological infiltration in an Instagram Reel.

Despite the moral inversion that social media seems to impel, real-world truth is still objective. Lies, which are endlessly abundant, create a kind of inflation that leads their purveyors to ruin. The demand for truth, which must be mined through hard persistence, is always there. We have to dig and dig.

The teaching of Jewish history and religion often converge on a single idea: Calamity may have been perpetrated by our enemies, but, ultimately, we bear responsibility. In this case, that means taking responsibility in the present. Our enemies were building their case over decades while we were resting on our laurels. The best time to have addressed this was then. The second-best time is now. *

ASHLEY RINDSBERG is the founder and CEO of NPOV. He was the runner-up to the 2025 DAO Grand Prize for investigative journalism and is the editor-at-large at *Pirate Wires*.