The Anatomy of Censorship and Propaganda: Evidence from Russian Wikipedias

Ekaterina Zhuravskaya, Ulrich Matter, Ruben Durante, Vladimir Avetian*

June 2026

Abstract

Authoritarian regimes shape citizens' knowledge and beliefs through censorship and propaganda. We study which narratives they suppress and promote using a rare natural experiment: Ruwiki, created in 2023 by cloning Russian-language Wikipedia and selectively rewriting its 1.9 million articles under regime-aligned editorial control. We compare the two platforms twenty months after the fork, using within-Wikipedia changes as a placebo benchmark. We find that Ruwiki's editing patterns resemble professional rather than volunteer-driven content production. An LLM-based comparison and machine-learning analyses of text deletions and insertions reveal a coherent ideology organized around three pillars: Russia's image vis-à-vis the rest of the world, domestic politics, and culture. The resulting narratives legitimize Russia's past and present acts of aggression, suppress accounts of domestic opposition and human-rights violations, and promote traditional values. Culture's centrality reveals how authoritarian information control penetrates domains far beyond formal politics, including morality, gender, family life, law, and psychology. Article deletions are rare and confined to topics that are difficult to reframe, while editing is broad, consistent with a preference for reframing existing content rather than leaving unanswered queries: 27% of articles undergo at least some change, 2.8% are heavily edited, and only 0.14% are deleted. Rewriting sensitive topics reduces readability, suggesting that ideological control is costly. An external benchmark using official high-school history textbooks confirms that Ruwiki's revisions move articles on modern history toward state-sanctioned narratives. The results show how authoritarian information control constructs a regime-aligned account of the world and Russia's place in it.

JEL Codes: P00, D83, L82, L86, D72, C55
Keywords: Censorship, Propaganda, Autocracy, Russia, Wikipedia, Text analysis, Large Language Models

---

*The authors are listed in *reverse alphabetical order*. All authors contributed equally to every part of the paper. Zhuravskaya: Paris School of Economics (EHESS) and CEPR, ekaterina.zhuravskaya@parisschoolofeconomics.eu. Matter: Bern University of Applied Sciences and University of St. Gallen, ulrich.matter@bfh.ch. Durante: National University of Singapore, ICREA-UPF, IZA, CESifo, and CEPR, rubendurante@gmail.com. Avetian: University Paris-Dauphine – PSL, vladimir.avetian@dauphine.psl.eu. We are grateful for helpful comments from Irena Grosfeld, Sergey Guriev, Jérôme Hergueux, and participants at the 17th TSE Digital Economics Conference (Toulouse, January 2025), the ECARES Political Economy Conference (Brussels, April 2025), the CEPR Workshop on Media, Technology, Politics, and Society (Milan, June 2025), and several invited seminars at various institutions. We thank Naman Agrawal, Gabriel Benevides, Simone Genetin, Aaron Schade, and Ren Jie Teh for outstanding research assistance. Ruben Durante gratefully acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (Grant Agreement No. 101125953).

I Introduction

Controlling the flow of information is a defining feature of authoritarian rule. Modern autocrats shape what citizens know and believe through two complementary instruments: censorship, which suppresses content damaging to the regime, and propaganda, which promotes narratives favorable to it. These instruments extend across every domain of knowledge, including politics, history, science, culture, religion, and morality, and operate through all the channels by which citizens form opinions, from news and school curricula to popular culture and the internet. The political economy of non-democracy has long studied *why* regimes control information: to preserve political power, reduce dissent, and mobilize citizens to undertake costly actions, such as fighting in a war or working for low pay (Besley and Prat, 2006; Edmond, 2013; Guriev and Treisman, 2019, 2020, 2022; Egorov and Sonin, 2024). Far less is known about *what* they control: which topics autocratic regimes prioritize when intervening in the information environment, and how the narratives they promote across topics fit together.

This question is central to understanding contemporary autocracies, but answering it is empirically challenging for two reasons. First, censorship is difficult to study because the suppressed content is, by definition, unobserved: researchers can observe what remains in the information environment, but rarely the counterfactual content that would have been available in the absence of censorship. Second, propaganda is difficult to detect at scale because establishing the accuracy, framing, and completeness of information is costly and often infeasible across many domains. As a result, it is difficult to distinguish systematic political manipulation from ordinary differences in editorial choices. The ideal empirical setting would therefore place politically manipulated content side by side, at scale, with an otherwise comparable but unbiased benchmark. Such a setting would allow researchers to identify not only the regime's preferred narratives, but also the precise domains in which politically constrained content departs from the counterfactual information environment, and to characterize the nature of these departures.

We exploit a unique empirical setting that approximates this ideal comparison. In the space of Russian-language online encyclopedias, two platforms coexist with starkly different editorial principles: Wikipedia and Ruwiki. Russian-language Wikipedia (ru.wikipedia.org) is a free, crowdsourced encyclopedia maintained by volunteers around the world under shared principles of neutrality and verifiability. Ruwiki (ruwiki.ru) is a state-sponsored "sovereign" alternative, created by cloning Russian-language Wikipedia in May 2023 onto Russian servers, selectively editing the cloned content, and officially launching the new version in January 2024 under a new brand name and interface.[1]

[1] Although Ruwiki's investors have not been publicly disclosed, an independent investigation links the project to structures associated with Yury Kovalchuk and Bank Rossiya. Kovalchuk is widely described as a close associate and friend of Vladimir

The emergence of Ruwiki, a deliberate clone of Russian-language Wikipedia subsequently rewritten under regime-aligned editorial control, offers a uniquely clean window into the regime's priorities in censorship and propaganda. When Ruwiki was "forked" from Wikipedia, the two encyclopedias contained identical text across nearly two million articles. Because Ruwiki preserved the same underlying MediaWiki infrastructure, we can trace full article histories. The shared origin and common technical structure allow us to compare the two platforms directly. Changes made within Ruwiki reveal how the cloned corpus was rewritten under political objectives and legal constraints, while contemporaneous changes on Wikipedia serve as a placebo benchmark for ordinary encyclopedic updating. The comparison therefore allows us to distinguish regime-aligned manipulation from routine revision and to identify which topics were altered and how. Overall, our findings reveal a deep and wide-ranging attempt to rewrite Russia's past and present across domains of knowledge, aligning politics, history, culture, morality, and foreign policy with a coherent, state-sanctioned view of Russia's identity and place in the world.

As of May 2026, Russian-language Wikipedia remains accessible in Russia, but it has faced persistent legal and regulatory pressure, including attempts to restrict its supply. Russian authorities have banned numerous Wikipedia articles, particularly those concerning the war in Ukraine and LGBTQIA+ issues; fined the Wikimedia Foundation, the nonprofit organization that hosts and operates Wikipedia and its sister projects, for non-compliance; and ordered search engines to label Wikipedia a violator of Russian law. Yet article-by-article supply-side censorship has proved technically difficult.[2] On the demand side, Russia's legal regime increasingly deters users from searching for or accessing prohibited content.[3] Ruwiki therefore offers users a legally compliant substitute: it preserves the appearance of an open online encyclopedia while ensuring that politically and legally sensitive content is removed or replaced with regime-compliant narratives.[4] This strategy relates closely to the porous censorship documented in China by Roberts (2018), who argues that information control need not eliminate access to disfavored information to be effective if it raises the cost of reaching it or diverts users toward more readily accessible alternatives.

The main goal of this paper is to describe systematically how content differs across the two platforms.

Putin, and Bank Rossiya as "Putin's wallet" (Novaya Gazeta Europe, 2025).

[2]See, for instance, Wikimedia Foundation (2022); Interfax (2022). Because Wikipedia uses HTTPS, blocking individual pages without restricting access to the platform more broadly is hard (see https://en.wikipedia.org/wiki/List_of_Wikipedia_pages_banned_in_Russia, accessed May 18, 2026).

[3]Individuals face potentially severe legal risks for "intentionally searching for or accessing extremist materials" online, with or without a VPN, including materials related to the so-called "LGBT movement," as designated by Russian authorities, and for disseminating, even in private communication, "false information about the Russian Armed Forces."

[4]After its launch, Ruwiki claimed to have opened editing to non-certified users until this option was revoked on January 1, 2026. However, the editing patterns we document below are inconsistent with substantial contributions by outsiders.

However, prior to analyzing content, we document that Ruwiki differs sharply from Russian-language Wikipedia in its mode of production. In particular, we examine editing activity on Wikipedia from 2019 to 2025 and on Ruwiki from its fork in May 2023 to 2025 to verify that they are consistent with Ruwiki being created by paid professionals working in a coordinated and directed fashion in contrast to Wikipedia, the product of independent volunteers. In the post-fork period, Ruwiki had roughly one-tenth of active monthly contributors compared to Wikipedia. Ruwiki contributors were much more active: at the peak, shortly after the fork, the average Ruwiki contributor changed roughly 126 pages of text per month. Editing intensity in Ruwiki subsequently declined, but by early 2025, the average contributor still changed 25 pages per month. By contrast, the average Wikipedia contributor's editing intensity remained stable at around 9 pages per month throughout the period from 2019 to early 2025. Ruwiki editing was also far more concentrated: the top 25 Ruwiki contributors account for 50% of all Ruwiki edits, and the top 200 account for 96%, compared with 23% and 54% on Wikipedia. Finally, Ruwiki editing closely follows the Russian work calendar: activity falls by over 50% on weekends, and by 62% on official holidays, compared to a normal working day, while Wikipedia editing does not follow this pattern. In addition, during workdays, Ruwiki editing intensity drops sharply right after Moscow-time office hours, while Wikipedia editing intensity remains stable until midnight.[5] Taken together, these patterns strongly suggest that Ruwiki is not simply a parallel volunteer community, but is more consistent with organized, schedule-driven editorial work.

Next, we ask how content differs across the two platforms. We approach this question at two levels of granularity. First, we characterize the broad narratives that Ruwiki suppresses and those it promotes, using a transparent pipeline combining statistical text analysis with large language models (LLMs). We then study how Ruwiki achieves this narrative shift by identifying the specific terms and phrases that characterize deleted and inserted text.

Because online encyclopedias record knowledge about the world at a given point in time, comparisons across platforms must hold the date, and therefore the underlying state of the world, constant. Our baseline comparison uses the January 1, 2025 snapshots. By then, Ruwiki had operated independently for twenty months, giving its editors time to decide what to delete, what to add, and what to leave unchanged. We therefore treat this date as capturing a new Ruwiki equilibrium: a regime-sponsored representation of the world. Since content on both platforms evolved after the fork, we also verify that the differences between the two platforms that we document on January 1, 2025 are driven primarily by edits made in Ruwiki,

[5] Trokhymovych et al. (2025) report related descriptive evidence on seasonal patterns in editing during the first four months after the fork.

rather than by changes made in Wikipedia after the fork.

Summarizing the narratives suppressed through censorship and promoted through propaganda is technically challenging in our setting because the two corpora contain about 1.9 million full-length articles each, cover a wide range of topics, and more than a quarter of the articles underwent at least some changes. We therefore begin by narrowing the focus. We restrict the analysis to articles that underwent *substantial* changes in Ruwiki relative to Wikipedia. This restriction is justified because many changes in Ruwiki are non-substantive, reflecting the new platform interface or cosmetic edits. In particular, we identify all articles deleted by Ruwiki editors and all heavily edited articles. We define an article as heavily edited if added text exceeds 10% of its original length, deleted text exceeds 10% of its original length, or the combined volume of added and deleted text exceeds 15% of its original length. Altogether, there are 56,835 such articles: 54,049 were heavily edited but remained in Ruwiki, while 2,786 were deleted completely.[6]

To identify broad narratives pursued through censorship and propaganda, we developed a three-step replicable pipeline that combines classical statistical association with LLM summarization. First, across the 56,835 heavily edited and deleted articles, we identify the bigrams and trigrams most strongly associated with additions and deletions, using χ^2 statistics on cleaned and lemmatized Russian text. This part of the analysis builds on phrase-based approaches to measuring media slant developed by Gentzkow and Shapiro (2010) and applies them to our setting. Second, we ask GPT-4.1, with temperature set to 0, to identify coherent narratives separately from the terms characteristic of deleted and inserted text. As substantive input, the model receives only the category label and the two term lists; it is not told the identities of the compared corpora, that the exercise concerns Russia, or that our hypotheses involve censorship and propaganda. Third, we verify the stability of the resulting narratives across levels of granularity, ranging from one to eight narratives per direction, and across 13 model and parameter configurations spanning six distinct LLMs: GPT-4.1, GPT-4o, Claude Sonnet 4.5, GPT-5, OpenAI o3, and DeepSeek-V3. Including DeepSeek-V3, a model produced by a Chinese company, helps alleviate the concern that the results reflect biases specific to Western LLMs.

The narratives identified by this procedure are coherent and remarkably stable across models and configurations. They reveal an editorial intervention organized around three thematic pillars. First, on *aggression, war, and foreign policy*, Ruwiki removes references to Russia's invasion of Ukraine, erases markers of Ukrainian sovereignty and territorial integrity, and suppresses Western and international perspectives on the conflict, while inserting content that legitimizes the annexation of Crimea and other

[6]Deleted articles mechanically fit our definition of heavily edited articles, since 100% of their text was deleted. We nevertheless distinguish outright deletion from heavy editing when examining how censorship and propaganda achieve their objectives.

Ukrainian territories and presents them as administratively integrated regions of the Russian Federation. Second, on *domestic politics and human rights*, Ruwiki suppresses discussion of opposition figures, as well as of human rights, civil liberties, independent media, and criticism of state institutions, while reinforcing the constitutional and legal authority of the Russian state. Third, on *culture and traditional values*, Ruwiki suppresses content related to LGBTQIA+ rights and Western liberal norms, while promoting traditional gender roles and presenting Russia as a defender of "traditional values" against what it portrays as Western, particularly European, moral decline. These narrative patterns constitute our first main finding.

A placebo that applies the same pipeline to changes within Wikipedia over the twenty months between the fork and January 2025 recovers a qualitatively different pattern. Additions broaden articles' historical, demographic, cultural, and sports-related context and add critical reception and recognition of achievements, while deletions primarily streamline them by removing redundant or outdated demographic and administrative information, bureaucratic references, duplicative population data, and minor biographical and local detail. Although the placebo captures factual updating related to major political and military events during the period, it does not reproduce the coherent directional pattern observed in the baseline. Instead, it reflects the ordinary expansion and maintenance of encyclopedic content. This placebo therefore strongly suggests that the three pillars reflect Ruwiki's editorial intervention rather than common temporal trends in Russian-language Wikipedia.

We then move to a more granular analysis, using machine learning and other complementary methods to examine the specific changes through which Ruwiki editors shift the corpus toward these narratives relative to its Wikipedia counterpart. We first examine outright deletions. The fact that heavily edited articles outnumber deleted articles by a factor of 19.4 is itself informative about the nature of censorship. Ruwiki editors appear to rely primarily on narrative alteration rather than outright removal. This is consistent with the idea that deleting sensitive topics entirely is costly: users who do not find information in Ruwiki can turn to other sources or to Wikipedia itself. For example, Ruwiki retains an article on Aleksei Navalny, but transforms its meaning: it foregrounds his inclusion on Russia's official list of terrorists and extremists and misrepresents his murder in a penal colony as the result of illness. At the same time, Ruwiki deletes articles about Navalny's corruption investigations, which are much harder to reframe.

As the number of deleted articles is relatively small, 2,786, we manually classify them into broad topics. This exercise yields four distinct clusters: (i) culture, consisting largely of pornography and LGBTQIA+ content, which together account for almost two-thirds of all deletions, but also including articles on heterosexuality, feminism, and religion; (ii) the war in Ukraine, particularly atrocities commit-

ted by the Russian army, such as the Bucha massacre; (iii) historical topics considered controversial by the Russian regime, including articles on the Soviet Union as an aggressor in 1939, the fall of the Russian Empire, the history of Ukraine before 1917, and antisemitism in the USSR; and (iv) domestic politics, including articles describing the activities of opposition and human-rights organizations.

In addition, we use an LLM to assign all 1.9 million articles in the corpus to 15 substantive, overlapping topic categories predefined by Wikipedia. We then compare the prevalence of these topics across samples. Deleted articles are much more concentrated in a small set of topics than the corpus as a whole and are particularly overrepresented in (i) *sex and gender*, (ii) *culture, media and entertainment*, (iii) *people/person*, and (iv) *law and order*.

We then use LASSO models based on bigrams and trigrams extracted from article titles and Wikipedia category labels (topic tags assigned to each article) to identify terms that robustly predict deletion. The terms consistently selected include *pornography article*, *LGBTQIA+*, *Russian invasion*, *invasion of Ukraine*, *war crime*, *Russian Federation*, and *caution editing*. The last term is Wikipedia's warning that an article is politically sensitive in Russia or Belarus and that editing it from within either country may expose contributors to risk; we henceforth refer to it as a contributor-risk notice. Because LASSO selects terms jointly, these predictors are markers of broader clusters of co-occurring concepts rather than an exhaustive list of the n-grams characterizing deleted articles.

Of the articles retained by Ruwiki, 522,135 (about 27% of the entire corpus) were edited, and 54,049 (about 2.8% of the entire corpus) were heavily edited. We first characterize the broad topics covered by these two groups. Heavy editing spans a much wider range of topics than outright deletion. Non-heavy editing is broader still, with a topic distribution closer to that of the corpus as a whole. In absolute terms, heavily edited articles are concentrated in categories that are also the largest in the whole corpus such as *history and politics*, *people/person*, *culture, media and entertainment*, and *geography*. Relative to their prevalence in the full corpus, however, the most overrepresented categories are (i) *culture, media and entertainment*, (ii) *history and politics*, (iii) *people/person*, (iv) *technology*, and (v) *law and order*. Smaller categories are especially revealing in relative terms: *sex and gender*, *psychology and psychiatry*, and *law and order* are two to three times as prevalent among heavily edited articles as in the full corpus. Thus, while outright deletion is concentrated in a narrow set of topics, heavy editing is a broader instrument: it reaches large domains such as history, politics, culture, and geography, and disproportionately targets smaller but ideologically salient areas related to gender, law, and psychology.

To understand how the reframing is implemented inside retained articles, we estimate a LASSO logit at the level of text segments, comparing phrases inserted into Ruwiki with phrases present in Wikipedia

but removed from Ruwiki, controlling for article fixed effects. The phrase-level results show that Ruwiki systematically deletes the language of Russian aggression and Ukrainian sovereignty, such as *Russian invasion of Ukraine*. These phrases are replaced with official, legalistic, and administrative terminology: *special military operation* and *military actions involving Ukraine*. Ruwiki also deletes mentions of prominent opposition, most prominently Aleksei Navalny, and LGBTQIA+ content, and explains what is *prohibited by the Russian Federation*. Category-specific estimates show the same logic across domains. In history and politics, war-related reframing dominates; in law and order, it is combined with the language of legal prohibition, foreign-agent designation, and regulation of gender and sexuality; in sex and gender, deletions remove the language of rights and recognition, while insertions introduce the language of pathology, regulation, and prohibition. We verify that changes to the same articles within Wikipedia between the fork and January 2025 do not exhibit the distinctive linguistic pattern. The results are robust to an alternative approach that uses Cramér's V to rank words and n-grams by how strongly they distinguish insertions from deletions, although it produces long lists of related terms corresponding to each of the concepts singled out by LASSO.

We then examine whether Ruwiki's interventions affect text quality. On average, heavily edited Ruwiki articles are substantially shorter than their Wikipedia counterparts, while differences in readability are small. This average, however, conceals striking heterogeneity. For particularly sensitive articles, identified by Wikipedia's contributor-risk notice, the Ruwiki versions are substantially denser, less lexically diverse, and markedly less readable than the corresponding Wikipedia articles. By contrast, edits to the same articles within Wikipedia over time produce no substantively meaningful change in text quality. Censorship and propaganda therefore impose a substantial readability cost.

Finally, as a reality check, we ground the encyclopedia comparison in an external benchmark: the two official 2023 Grade-11 history textbooks used in the final year of Russian high school, widely understood as reflecting the regime's preferred historical narrative. We measure whether each encyclopedia version's account of matched postwar and contemporary-history topics contradicts the state-sanctioned textbook account. We match heavily edited encyclopedia articles to textbook subchapters and retain the top 100 article–subchapter pairs by semantic similarity. Wikipedia articles are 13 percentage points more likely to contain at least one contradiction with the textbook benchmark than their Ruwiki counterparts and have about three times as many contradictions as their Ruwiki pair. Again, the within-Wikipedia placebo for the same articles shows no significant change in their alignment with the textbook content. Ruwiki's editorial intervention thus moves content toward state-sanctioned historical narratives, as one would expect.

Overall, these analyses yield four important lessons. First, culture is an integral part of the regime's

censorship and propaganda strategy. It is one of the three main priorities of Ruwiki's intervention, alongside Russia's image vis-à-vis the rest of the world and domestic politics. Two related mechanisms may explain the prominence of culture. Traditional-values narratives may cultivate obedience and respect for authority and hierarchy not only within families but also, by extension, in political life. The connection between authority within the household and the polity has long been emphasized by political thinkers, from Aristotle (ca. 350 BCE) to Mill (1869). At the same time, the regime presents itself as a defender of "traditional values" against an allegedly decadent West, with Europe increasingly serving as its principal cultural foil.[7] The second lesson concerns how, rather than what, the regime seeks to control. Outright censorship without narrative replacement is limited. The regime appears to prefer retaining potentially sensitive topics while changing the interpretive frame through which they are presented rather than deleting them altogether. One reason is that deletion may be politically costly: users who do not find information on Ruwiki may turn to other sources. Third, censorship and propaganda extend across a wide range of domains. Ruwiki alters content related to Russia's foreign policy, domestic politics, and culture not only in articles directly concerning politics, history, culture, and geography, but also in seemingly distant fields such as nature, science, technology, philosophy, and sports. Finally, censorship and propaganda come at the cost of lower-quality text. Ruwiki often replaces standard encyclopedic descriptions with official, bureaucratic, or legalistic terminology and euphemisms, making the resulting articles not only less informative but also less readable. This pattern echoes classic accounts of propaganda as a pervasive linguistic environment in which official formulas and repeated vocabulary naturalize the regime's worldview (Ellul, 1965; Klemperer, 2000; Yurchak, 2006).

Taken together, our results offer what is, to our knowledge, the first quantitative anatomy of an authoritarian regime's priorities in censorship and propaganda. The paper contributes to several strands of social science literature. First, we contribute to the literature on information control by modern dictators, which models how autocrats strategically manage information to sustain power, preserve popularity, and avoid accountability (Besley and Prat, 2006; Edmond, 2013; Gehlbach and Sonin, 2014; Shadmehr and Bernhardt, 2015; Guriev and Treisman, 2019, 2020, 2022; Boleslavsky, Shadmehr and Sonin, 2021; Egorov and Sonin, 2024). Second, we speak to the empirical literature on censorship in authoritarian regimes, primarily focused on China, which documents how governments suppress politically sensitive content while allowing information that does not threaten regime stability (King, Pan and Roberts, 2013,

[7] For example, in his address to the Valdai International Discussion Club on September 19, 2013, Vladimir Putin contrasted Russia's commitment to traditional and Christian values with what he described as their abandonment in many "Euro-Atlantic countries" (https://web.archive.org/web/20220327135245/http://en.kremlin.ru/events/president/news/19243, accessed June 28, 2026).

2014; Lorentzen, 2014; Qin, Strömberg and Wu, 2017; Roberts, 2018; Chen and Yang, 2019). Third, we contribute to research on propaganda strategies in authoritarian contexts, which studies how regimes shape narratives through state media, coordinated messaging, and selective framing (Lu and Pan, 2021; Mattingly and Yao, 2022; Stukal et al., 2022; Pan, Shao and Xu, 2022). Fourth, we relate to the large empirical literature on the political and non-political effects of state-controlled mass media (Enikolopov, Petrova and Zhuravskaya, 2011; Yanagizawa-Drott, 2014; Adena et al., 2015; Durante, Pinotti and Tesei, 2019; Blouin and Mukand, 2019); see also the surveys by DellaVigna and Gentzkow (2010); Enikolopov and Petrova (2015); DellaVigna and La Ferrara (2015).

Our contribution relative to these literatures is to observe censorship and propaganda jointly and across the full encyclopedic domain. Existing work typically studies particular media outlets, platforms, topics, or episodes. By contrast, the Ruwiki fork allows us to compare two comprehensive corpora that initially shared the same content and then were subsequently edited under different institutional control. This setting allows us to conduct a quantitative inventory of the topics an authoritarian regime targets, the narratives it suppresses and promotes, and its choice between deleting information and reframing it.

Methodologically, we contribute to the growing use of large language models for narrative analysis in social science. Our narrative-extraction pipeline is transparent, replicable, and blind to the identities of the two corpora. Robustness across models and the within-Wikipedia placebo comparison limit concerns that the results are driven by model-specific political priors or politically shaped training data (Waight et al., 2026). Our approach provides an LLM-based alternative to the structured narrative-extraction method developed by Ash, Gauthier and Widmer (2024). Our use of LLMs as a measurement tool is supported by recent evidence that GPT can code qualitative data accurately, with results robust to semantically equivalent prompts and often comparable to human coding (Asirvatham, Mokski and Shleifer, 2026).

The closest study in terms of setting is independent concurrent work in computer science by Trokhymovych et al. (2025), who document descriptive patterns during Ruwiki's first four months after the fork, from May to September 2023, and use machine learning to classify early changes across 33,664 articles. By contrast, we study the platform twenty months after the fork, once its editorial operation had matured and the two corpora had diverged on 522,135 articles. Our corpus-level methodology recovers the broad ideological narratives that organize the intervention across a wide set of domains. We also analyze the strategic choice between deletion and narrative replacement, quantify the readability cost of manipulation, and validate the direction of historical reframing against official history textbooks.

The rest of the paper is structured as follows. Section II describes the institutional background. Section III presents the data collection. Section IV documents the editing-activity differences between

Wikipedia and Ruwiki. Section V identifies the coherent narratives that Ruwiki edits promote and suppress. Section VI examines in detail which articles are deleted and which are heavily edited. Section VII benchmarks the two encyclopedias against official Russian history textbooks. Section VIII concludes.

II Background

Russian-language Wikipedia has long been one of the largest and most visited online reference resources in Russia. In March 2026, it contained more than two million articles, was maintained by roughly 10,000 active monthly contributors, and attracted about 252 million monthly visits.[8] Its volunteer-based governance structure rests on the principles of neutrality, verifiability, and reliable sourcing, with independent administrators selected through community consensus playing a central role. These features have made it difficult for any single political actor, including the Russian state, to capture the platform, despite evidence that political actors have attempted to influence contested content (Greenstein and Zhu, 2012; Potapenko and Kamionka, 2023; Steinsson, 2024). For Russian users, this has made Wikipedia one of the few remaining high-traffic websites where contested political and historical topics can be accessed outside direct state control.

The Russian state has repeatedly attempted to bring Wikipedia content into line with official narratives. Pro-government editing efforts targeted articles on politically sensitive topics, such as the downing of Malaysia Airlines Flight 17, the annexation of Crimea, Russian military activity in Syria, and the 2022 invasion of Ukraine. Such edits were typically reversed by the broader community of volunteer editors when they violated Wikipedia's neutrality and sourcing rules.

The conflict escalated sharply after Russia's full-scale invasion of Ukraine in February 2022. Wikipedia's coverage of the war used the term "invasion," documented civilian casualties and allegations of war crimes, and directly contradicted the Kremlin's official euphemism, "special military operation." Roskomnadzor repeatedly demanded the removal or alteration of Wikipedia articles on the war, and the Wikimedia Foundation was fined for non-compliance. In July 2022, Russian authorities also ordered search engines to warn users that the Wikimedia Foundation was violating Russian law because Wikipedia continued to host prohibited content (Wikimedia Foundation, 2022; Interfax, 2022). Senior officials criticized Wikipedia publicly, but the regime remained reluctant to block it outright while no comparable domestic alternative existed (Fadeyev, 2023; Shadayev, 2023).

[8] Article counts and active-user counts are from Russian Wikipedia's public statistics page, https://ru.wikipedia.org/wiki/Special:Statistics; page-view statistics are from Wikimedia Statistics, https://stats.wikimedia.org/#/ru.wikipedia.org/reading/total-page-views/normal|bar|1-year|~total|monthly (both accessed May 27, 2026).

Ruwiki emerged as that domestic alternative.[9] In May 2023, the content of Russian-language Wikipedia, including articles, categories, revision histories, internal links, and the underlying MediaWiki infrastructure, was copied onto Russian servers in a software fork. The cloned content was then selectively rewritten and made publicly available at www.ruwiki.ru on January 15, 2024 as an official launch. (The beta version was launched in June 2023.) The project was led by Vladimir Medeyko, the former director of Wikimedia Russia. After launching Ruwiki's beta version in June 2023, Medeyko was stripped of his position at Wikimedia Russia by the organization's members and later banned indefinitely from editing Wikimedia sites by the Wikimedia Foundation (The Signpost, 2024). Wikimedia Russia itself subsequently ceased to operate; its chapter agreement with the Wikimedia Foundation was terminated effective March 27, 2024 (Wikimedia Foundation Meta-Wiki, 2026).

As mentioned in the introduction, Ruwiki is hosted in Russia, financed by undisclosed investors, and widely understood to be aligned with the Kremlin's agenda. Independent investigations link the project to structures associated with Yury Kovalchuk, a billionaire media mogul, majority shareholder of Bank Rossiya, and a close ally of Vladimir Putin (Novaya Gazeta Europe, 2025). Through National Media Group and related entities, Kovalchuk also controls several federal television channels, the newspaper *Izvestia*, and, since 2021, VKontakte, Russia's largest social-media platform. Although Ruwiki's audience remains far smaller than Wikipedia's, the platform has been promoted through an extensive advertising campaign, including outdoor billboards and ads in the Moscow Metro (Wikipedia Signpost, 2025). According to Ruwiki CEO Vladimir Medeyko, the platform's monthly audience reached 13 million users in 2025 (Tsvetkova, 2025). Although Ruwiki initially presented itself as open to outside contributors, as of January 1, 2026, it officially restricted editing to "verified users, with preference given to authors with relevant education and experience working with information" (Kommersant, 2026). Whether editing had already been fully controlled de facto remains unknown. Senior Russian officials have suggested that the existence of a domestic analogue removes the main obstacle to a future ban on Wikipedia (Khinshtein, 2023).[10]

[9] There were several earlier attempts to create regime-friendly Russian-language online encyclopedias alternative to Wikipedia, including Runiversalis and Znanie, but all of them failed to produce a comparable substitute (Wikipedia Signpost, 2025).

[10] After Russia's 2022 invasion of Ukraine, the Russian state banned Facebook and Instagram (designating Meta "extremist"). YouTube was heavily throttled in Russia, while TikTok suspended new content out of self-censorship in response to the risks associated with spreading information considered "fake news" by the Russian law.

III Main Data

Our data collection exploits the fact that Ruwiki forked not only the content of Russian-language Wikipedia, but also its MediaWiki technical infrastructure. At the time of the fork, the two platforms shared the same articles, categories, internal links, and revision histories. Both platforms also store articles in raw Wikitext and expose comparable MediaWiki APIs through which article content, revision histories, and contributor metadata can be queried. This common origin and shared infrastructure allow us to compare the two encyclopedias systematically and to distinguish subsequent Ruwiki rewriting from ordinary Wikipedia updating.

Collection. Unlike Wikipedia, which makes downloadable database dumps of its full edit history available, Ruwiki does not publish data dumps. We therefore collect Ruwiki data through the MediaWiki API. For Wikipedia, we use both the API and the public dumps available at https://dumps.wikimedia.org/, accessed May 27, 2026. Ruwiki's back-end infrastructure was fully functional and publicly reachable throughout our data-collection period, and we did not detect any analytically relevant divergence from the standard MediaWiki structure.[11]

Sample. Our analysis is restricted to the *main encyclopedic namespace*, ordinary encyclopedia articles, rather than discussion pages, user pages, or administrative content, and to articles that existed on Wikipedia at the time of the fork on May 1, 2023 or were created between the fork and January 1, 2025. This restriction captures the universe of content that Ruwiki could, in principle, have inherited, modified, or created. The resulting corpus contains about 1.9 million main-namespace articles with category labels, organizational tags that editors attach to articles to group them with similar subjects.[12]

Editing activity, article versions, metadata. We collected complete article-edit histories through February 1, 2025, covering both the pre- and post-fork periods for Wikipedia and the post-fork period for Ruwiki. For each edit, we obtained the precise timestamp, the username or IP address of the contributing editor, the full Wikitext content of the article immediately after the edit, and the additional metadata fields described in Online Appendix Table A1. The resulting database includes the version of each article on Wikipedia immediately before the fork, as well as subsequent versions of those same articles after they

[11] The data-collection pipeline, including a Rust-based parser, AWS-based parallel processing, and quality-assurance procedures, is detailed in Online Appendix A.

[12] The small set of articles without category labels, typically very recent stub pages, is excluded from analyses that require topical classification but included in analyses comparing Wikipedia and Ruwiki content.

were copied to Ruwiki and modified by Ruwiki editors. In total, the dataset comprises approximately 79.5 million article edits across 3.8 million main namespace articles in both platforms: 78.26 million edits on Wikipedia between November 2002 and February 2025 and 1.25 million on Ruwiki between the fork and February 2025. Wikipedia edits span the history of the platform, but we focus on the period from January 2019, which covers 27.28 million Wikipedia edits.[13] For each article version, we also collect the article's original Wikipedia categories. A typical article carries several such categories, which provide a concise summary of its substantive content.[14]

Comparison design. Our primary comparison contrasts each article's content on Wikipedia and on Ruwiki on the same date, January 1, 2025, twenty months after the fork. This snapshot captures the cumulative effect of all edits made on each platform since the fork in May 2023, while abstracting from edit-by-edit dynamics. Differences between the two platforms at this date reflect two forces: selective editing by Ruwiki's editorial team and the natural evolution of Wikipedia content over the same period. To assess the second channel, we compare Wikipedia articles at the fork to their versions on January 1, 2025. This within-Wikipedia placebo comparison allows us to distinguish ordinary encyclopedic updating from Ruwiki's editorial intervention. For completeness, we also conduct the analogous within-Ruwiki comparison, tracking how the cloned articles changed between the fork and January 1, 2025.

Deleted articles. In the baseline comparison, an article is classified as deleted if it belonged to the main encyclopedic namespace at the time of the fork, was absent from Ruwiki on January 1, 2025, and remained present on Wikipedia on that date. We exploit the fact that each Wikipedia page has a unique numerical identifier, which Ruwiki retained when it cloned the corpus. We first identify all main-namespace articles with active revision histories on Wikipedia at the fork and then check whether each article is present, that is, has an active revision history, on Ruwiki and on Wikipedia as of the comparison date. For articles with no Ruwiki revision activity, we further verify absence through direct API queries and keyword searches.[15] In total, 2,786 articles meet these criteria in the baseline comparison on January 1, 2025.

[13] The number of articles "ever edited" exceeds the number of articles in the main encyclopedic namespace because Wikipedia and Ruwiki include administrative, talk, user-page, and other non-encyclopedic namespaces.

[14] Editors assign categories by adding Wikipedia markup code, for example `[[Category:History]]`, to the article's source text, typically near the end of the article. The MediaWiki system then automatically adds the article to that category's listing, creating hierarchical taxonomies that allow readers to browse related content. Categories appear as clickable links at the bottom of articles. For example, a biography of a Russian opposition politician might carry categories such as *politicians of Russia*, *Russian dissidents*, and *20th-century Russian male writers*. See https://en.wikipedia.org/wiki/Help:Category (accessed June 28, 2026) for details.

[15] For the small number of articles that were deleted on Ruwiki and later re-created with new identifiers, we perform supplementary matching by article title to avoid false positives.

Heavily edited articles. For articles that exist on both platforms, we use a custom Rust-based Wikitext parser (Portström, 2018) to extract the readable plain-text content of each article, stripping away templates, infoboxes, references, tables, and other formatting markup. We then apply the sequence-matching algorithm developed by Ratcliff and Metzener (1988) to identify text segments that appear in one version but not the other: insertions into Ruwiki and deletions relative to Wikipedia. This procedure identifies 522,135 articles with at least some textual modification across platforms in the baseline January 2025 comparison. We define an article as heavily edited if deleted text exceeds 10% of its original length, inserted text exceeds 10% of its original length, or the combined volume of inserted and deleted text exceeds 15% of its original length. We further restrict attention to articles with at least 100 words in the original Wikipedia version. This is a minor restriction: fewer than 1% of main-namespace Wikipedia articles are shorter than 100 words. The resulting set contains 54,049 heavily edited articles. The thresholds exclude both very short stub articles, where small absolute changes can produce misleadingly large relative measures, and articles with only minor stylistic edits, focusing the analysis on substantive content modifications. Online Appendix A.2 documents the parsing, identification of text additions and deletions, and threshold construction in detail. Online Appendix A.3 provides details on the data cleaning procedure.

Topic classification. To analyze the thematic structure of editorial intervention, we classify all 1.9 million main-namespace Wikipedia articles into broad substantive topics using an LLM. The inputs to the classifier are each article's title and its original Wikipedia category metadata. We refer to the resulting output as LLM categories throughout the paper. Our LLM categories consist of fifteen substantive topics: history & politics, culture/media & entertainment, sex & gender, law & order, people/person, environment, religion, science, language, philosophy, psychology & psychiatry, technology, sports, nature, and geography. This list is adapted from Wikipedia's own list of controversial topic areas, where editorial disputes have historically been common (Wikipedia, 2024), with minor expansions to cover parts of the encyclopedic corpus that are not central to disputes. Classification is performed by the DeepSeek-V3 large language model (DeepSeek-AI et al., 2024). The model was asked to assign articles to seventeen binary categories, with history coded separately from politics and culture/entertainment coded separately from media. Articles can belong to multiple categories simultaneously: a biography of a historical political figure, for example, may be classified as *history*, *politics*, and *people/person*. Because of substantial overlaps, we combine history and politics into a single category; and, similarly, culture & entertainment is combined with media into another single category. Online Appendix B reports the full classification prompt, processing details, and category-overlap structure.

For the 2,786 deleted articles, we also conduct a manual topic classification using two human annota-

tors, both of whom are authors of the paper. Unlike the LLM-based classification, which assigns articles to overlapping substantive categories, the manual classification assigns each deleted article to exactly one of 25 fine-grained topics. These topics were defined inductively by the annotators and capture the apparent reason for deletion. The coding task was not blind: annotators knew that the articles had been deleted from Ruwiki. The two authors independently reviewed each article's title, original Wikipedia categories, and, when necessary, a snapshot of the pre-deletion content, and assigned each article to a substantive topic.[16] We then grouped the 25 manually coded topics into five umbrella categories: culture, history, war/geopolitics, domestic politics, and other. In Section VI.A below, we report the complete mapping between the manual topic classification and LLM-categories of deleted articles.

Contributor-risk notices. Wikipedia attaches various page-specific warnings, known as *editnotices*, to articles requiring particular care. We focus on the "Caution Editing" notice, which warns contributors that editing the article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. We use the presence of this notice as an externally supplied marker of topics that are especially sensitive to the Russian regime. This contributor-risk notice appears on 3.16% of articles deleted from Ruwiki, 2.49% of heavily edited articles, 0.53% of articles edited but not heavily, and only 0.05% of unedited articles. Below, we examine whether Ruwiki's interventions have different effects on text quality for articles carrying this notice.

Official history textbooks. To benchmark encyclopedia narratives against an external state-sanctioned source, we use official Russian high-school history textbooks. We focus on the two 2023 textbooks used in the final year of high school (Grade-11): one on Russian history (Medinsky and Torkunov, 2023) and one on world history (Medinsky and Chubar'yan, 2023). Both were prepared under the supervision of Vladimir Medinsky, a presidential aide to Vladimir Putin and former Minister of Culture. These textbooks were introduced as part of a broader post-invasion effort to rewrite and standardize history instruction in line with the regime's official interpretation of Russia's war against Ukraine and its place in Russian and world history (BBC Russian, 2023; Deutsche Welle, 2023; Foreign Policy, 2023). They therefore provide a clear benchmark for the regime's preferred historical narrative, designed for transmission to teenage students.

[16] Disagreements between coders, about 5% of cases and almost entirely concerning whether a marginal article belonged to "culture" or to "other," were resolved by consensus during a joint review.

IV Editing Activity: Ruwiki vs. Wikipedia

Although Ruwiki was publicly presented as a Russian-language analogue of Wikipedia, the editing patterns of the two platforms are difficult to reconcile with the same crowdsourced production model. Even before January 1, 2026, the date when Ruwiki officially closed editing to outside contributors, its activity was highly concentrated among a small number of editors and displayed systematic time-use patterns consistent with coordinated, professional editing. By contrast, Russian-language Wikipedia exhibits the more diffuse and dispersed activity associated with volunteer production. Panels A–E of Figure 1 present five facts that support this interpretation.

Number of contributors. Panel A plots the number of distinct users, who made at least one edit in a calendar month between January 2019 and our latest snapshot, February 2025. Before Russia's full-scale invasion of Ukraine in February 2022, active editing on Russian Wikipedia was stable, averaging 9.6 thousand monthly contributors. After February 2022 and before the May 2023 fork, the number of active contributors declined by about 40 users per month ($p = 0.017$).[17] The change in slope at the onset of the war is marginally significant ($p = 0.088$ with Newey-West (NW) standard errors with two lags). After the fork, the number of monthly Wikipedia contributors fell further, by about 856 users on average relative to the extrapolated pre-fork trend (this fall is statistically significant, $p < 0.001$ with NW standard errors) and continued to decline with the same rate as before. By February 2025, Wikipedia had 7,362 active contributors.

Ruwiki, in contrast, had far fewer active editors. In the post-fork period, the number of monthly active Ruwiki editors averaged 731, roughly one-tenth of the corresponding number on Russian Wikipedia. This number was stable over the observation period, with two spikes: 1,624 contributors in the month between the fork and the release of the beta version and 1,079 contributors in the month before the full public launch.

As combined activity on Wikipedia and Ruwiki after the fork is statistically indistinguishable from the extrapolated trend in Wikipedia contributors predicted from the period between the start of the war and the fork ($p = 0.50$), the observed pattern is consistent with substitution of editing activity from Wikipedia to Ruwiki. This pattern, however, does not mean that those editors who might have moved from Wikipedia to Ruwiki were independent of the Russian state. Russian agents attempted to manipulate Wikipedia

[17] Possibly, this occurred out of fear of persecution in Russia or Belarus, where such precedents have occurred; Wikipedia has consequently marked Russian-language Wikipedia articles on particularly sensitive topics with a contributor-risk notice warning users not to reveal their identity when editing (Радыё Свабода, 2025; Reform.news, 2026).

articles in ways consistent with pro-Kremlin narratives on many occasions and were consistently blocked by the Wikipedia editors (see, e.g., Miller et al., 2022; Kurek, Budak and Gilbert, 2025; EUvsDisinfo, 2026).

Per-contributor edits: size, number, and concentration. There is a sharp contrast between the two platforms in both the average level and the concentration of editorial activity. We document this using 1,247,135 Ruwiki edits and 7,985,350 Wikipedia edits from the post-fork period, May 2023 to February 2025, as well as 19,292,169 Wikipedia edits from the pre-fork period, January 2019 to April 2023. For each edit, we observe the contributor's username and proxy the size of the edit by the absolute value of the change in article length before and after the edit.[18]

Ruwiki contributors make more changes on average. As shown in Panel B of Figure 1, the size of monthly edits, measured in kilobytes deleted or added, is significantly higher for Ruwiki contributors. Wikipedia contributors make, on average, 17KB of changes per month (with a 95% confidence interval (CI) of [16.7; 17.4]). This level of activity is very stable over time. Ruwiki contributors were much more active at the beginning: two months after the fork, in July 2023, average monthly edit size per contributor reached 237KB (95% CI: [65; 409]). It then gradually declined to 50KB in February 2025 (95% CI: [30; 70]). Thus, at the peak of activity, the average Ruwiki editor changed roughly 126 pages of text per month; by early 2025, this workload had declined to about 25 pages per month. By comparison, the average Wikipedia contributor changes about 9 pages of text per month. These differences are both economically and statistically significant.

Online Appendix C shows that these differences are driven by both a larger number of edits and larger changes per edit: Ruwiki editors make more edits per month and, on average, make larger changes in each edit than their counterparts on Wikipedia (Panels A and B of Appendix Figure A4).[19] In addition, Ruwiki contributors make more edits that delete text and more edits that add text than Wikipedia contributors (Panels C–E of Appendix Figure A4). In the first four months after the fork, the average Ruwiki contributor deleted more than twice as much text as they added, as measured by aggregated net deletions and net additions across edits. In the first half of 2024, deletions and additions were similar in magnitude. In the second half of 2024, the pattern reversed: contributors added roughly twice as much text as they deleted.

[18]This is only a rough proxy for edit size, as we observe only the net change in length per edit. In the analysis of article content below, we instead compare the actual text added and the actual text deleted.

[19]The number of edits per Ruwiki contributor increases over time, consistent with public reports that Ruwiki editors actively use AI to generate and edit content (e.g., Rozhkov, 2025). Ruwiki editors have also reportedly added offline encyclopedic entries in many domains based on Russia's official offline encyclopedias (e.g., Khimshiashvili and Kanaev, 2024). In our analysis of the patterns of editing activity, we focus on pages that existed in Wikipedia at the time of the fork and therefore formed the starting point of Ruwiki.

In Wikipedia, by contrast, both monthly additions and monthly deletions per contributor remained stable. Over the full post-fork period, Ruwiki contributors deleted, on average, 6.6 times more text and added 2.5 times more text than Wikipedia contributors.

The concentration of editing activity among contributors is also much higher in Ruwiki than in Wikipedia. Over the entire post-fork period, the probability that two randomly selected edits were made by the same contributor was 3.7 times higher in Ruwiki than in Wikipedia (0.0177 in Ruwiki vs. 0.0048 in Wikipedia). Panel C of Figure 1 presents cumulative concentration curves for the number of edits on the two platforms, ranking top 200 contributors from most to least active. The top 25 contributors account for 50.1% of all Ruwiki edits, compared with 23.4% on Wikipedia. The skew extends far into the right tail: the top 100 Ruwiki contributors account for 89.4% of all edits on the platform, compared with 42.1% on Wikipedia; the top 200 account for 96.3% on Ruwiki, compared with 54.3% on Wikipedia.

The timing of edits. The temporal structure of editing distinguishes the two platforms even more sharply. We first examine how editing activity of top 200 contributors varies by type of day: ordinary weekdays, weekends, and official Russian holidays. We estimate a Poisson pseudo-maximum-likelihood (PPML) model, which is well suited for count outcomes with zeros, regressing the daily number of edits per contributor on indicators for days of the week and holidays, their interactions with the Ruwiki platform, and contributor fixed effects. The omitted category is an ordinary Monday. From these regressions, we recover the average percentage change in the number of edits per contributor for Ruwiki and Wikipedia, and plot these coefficients, together with their 95% confidence intervals, in Panel D of Figure 1. Editing activity is stable across ordinary weekdays on both platforms. However, Ruwiki activity falls sharply on weekends and holidays: relative to an ordinary Monday, the number of edits per contributor is 48.2% lower on Saturdays, 57.1% lower on Sundays, and 61.5% lower on holidays. This pattern contrasts sharply with Wikipedia. There is no significant difference in the number of edits per Wikipedia contributor between an ordinary Monday and either Sunday or a holiday. On Saturdays, Wikipedia contributors actually significantly increase the number of edits by 10.9% relative to Monday.

We also examine the pattern of activity of top contributors within an ordinary working day. Using PPML, we regress the number of hourly edits per contributor on indicators for hours of the day, measured in Moscow time, with 15:00–16:00 as the omitted category. We restrict the sample to ordinary workdays and control for contributor fixed effects. Panel E of Figure 1 reports the estimated percentage changes in the number of edits per contributor-hour relative to 15:00–16:00, separately for the top 200 contributors in Ruwiki and Wikipedia, together with 95% confidence intervals. The timing of editing activity is remarkably similar across the two platforms during standard office hours, from 9:00 to 18:00, and during

the late-night period, from 3:00 to 9:00. The two platforms diverge sharply, however, after the end of the working day. Ruwiki contributors' editing activity drops significantly after office hours: relative to 15:00–16:00, the number of edits falls by 18.1% by 19:00, 44.9% by 20:00, and 56.9% by 21:00. By contrast, Wikipedia contributors continue editing at roughly the same rate as in the afternoon until midnight; their activity begins to decline significantly only after midnight.

Together, the day-of-week and within-day patterns in Ruwiki editing closely mirror the Russian working-day schedule and official holiday calendar. This is also evident in the raw-data distributions, presented in Appendix Figure A5. Trokhymovych et al. (2025) present a similar descriptive pattern for the editing activity during the first four months after the fork.

Overall, we document that Ruwiki editing differs systematically from the volunteer editing model observed in Wikipedia. Ruwiki activity is concentrated among a small group of contributors, per-contributor output is substantially higher, activity is more concentrated on weekdays and during Moscow working hours, and editing falls sharply on official Russian holidays. Wikipedia editing is more dispersed, has lower per-contributor output, and does not follow Moscow office hours or the Russian work calendar. This difference in patterns strongly suggests organized, schedule-driven editing by a small core of paid professional editors generating the bulk of editorial activity in Ruwiki, in contrast to decentralized spare-time volunteer activity in Wikipedia.

This interpretation matters for the content analysis below. It suggests that differences between Ruwiki and Wikipedia should not be interpreted simply as the organic divergence of two volunteer communities. The next section examines these content differences.

V What Narratives Does Ruwiki Suppress and Promote?

Which coherent narratives does Ruwiki's editorial intervention suppress, and which does it promote? In this section, we identify these narratives relative to Wikipedia using a transparent and reproducible procedure. Three considerations guide our methodology. First, the procedure must use the full set of editorial changes across all deleted and heavily edited articles rather than a hand-picked sample. Second, it must recover semantic structure from large lists of Russian-language terms without requiring the analyst to read each article. Third, it must synthesize statistical patterns into interpretable narrative themes while minimizing the influence of the model's and analyst's priors. We meet these requirements with a three-step pipeline that combines classical statistical association with controlled LLM summarization.

V.A Method

Step 1: Statistical signal discovery. For each of the 56,835 deleted or heavily edited articles, comprising 54,049 heavily edited articles and 2,786 deleted articles, we extract cleaned and lemmatized text segments that are unique to one version: insertions, which appear in Ruwiki but not in Wikipedia, and deletions, which appear in Wikipedia but not in Ruwiki. We then pool these segments across articles and construct a document–term matrix of bigrams and trigrams.[20] For each n-gram, we construct a 2×2 contingency table that classifies text segments by whether they contain the n-gram and whether they are insertions or deletions. We then compute the Pearson χ^2 statistic for independence. Because the table margins reflect the total numbers of insertion and deletion segments, the statistic tests whether the share of segments containing the n-gram is the same among insertions and deletions, accounting for the difference in their total volumes. The sign of the association indicates whether an n-gram is diagnostic of insertions or deletions, and we rank n-grams by the χ^2 statistic separately in each direction. This step is purely descriptive.

Step 2: LLM-based narrative extraction. We pass the top 500 n-grams in each direction to GPT-4.1, at temperature 0, for replicability, and ask the model to identify the coherent narratives implied by each list. The full prompt is reproduced in Online Appendix D. Two design choices are central. First, the prompt is blind to the source and purpose of the comparison: it does not mention Russia, Ruwiki, censorship, or propaganda, nor does it identify the two corpora being compared. The model is told only that the two lists contain terms statistically associated with additions to and deletions from Wikipedia articles within a given category and is asked to identify potential narratives reflected in these patterns. As substantive input, it receives only the category label and the two term lists. This blinding is designed to prevent the model from importing prior beliefs about Ruwiki or being led by the framing of the exercise toward a predetermined conclusion about the Russian regime's censorship and propaganda. Second, the prompt requests a fixed number of narratives, i.e., granularity. Our preferred specification asks the model to identify seven deletion narratives and seven insertion narratives.

[20] We exclude unigrams from the main analysis because multi-word phrases such as "human rights," "special military operation," or "LGBT propaganda" carry semantic content that isolated words often do not. We retain only n-grams that appear in at least 5 but no more than 5,000 distinct text segments, approximately 0.6 percent of all pooled insertion and deletion segments, and in at least 4 distinct articles. The full preprocessing pipeline, including spaCy and pymorphy2 lemmatization, stop-word filtering, removal of Wikipedia-specific markup remnants, and part-of-speech filtering to keep only n-grams containing at least one noun, is detailed in Online Appendix A.3.

Step 3: Stability across granularity and models. To assess sensitivity to this choice of granularity, we rerun the narrative-extraction pipeline separately for deletions and insertions, varying the requested number of narratives from one to eight. A coherent thematic structure would be hierarchical: as the requested number increases, broad narratives would subdivide into more specific ones rather than be replaced by an unrelated set of themes. For each pair of adjacent granularity levels, we ask GPT-4.1, with temperature set to 0, to classify the relationships between narratives at the coarser and finer levels, which we summarize as *still present*, *split*, or *new*. The first two classifications preserve the underlying hierarchy, whereas *new* indicates that a narrative has no counterpart at the preceding, coarser level. Online Appendix Section D.3 describes the procedure in detail.

We further re-run the entire pipeline in an independent replication of the baseline specification and under 12 alternative model and parameter configurations: GPT-4.1, GPT-4o, and Claude Sonnet 4.5 each at temperatures 0, 0.5, and 1; OpenAI o3 with medium reasoning effort; GPT-5 with low and high reasoning effort; and DeepSeek-V3 (DeepSeek-AI et al., 2024). Developed by a Chinese AI company, DeepSeek-V3 helps address concerns that the results may reflect systematic biases shared by Western commercial LLMs.

Within-Wikipedia placebo, narratives. A potential concern is that the cross-platform comparison conflates two channels: Ruwiki's editorial intervention, which is the channel of interest, and changes in Wikipedia over the same twenty-month period that were not incorporated into Ruwiki. Online Appendix Section E develops the conceptual framework underlying these comparisons. To distinguish the two channels, we apply the same pipeline, including n-gram extraction, χ^2 ranking, LLM prompt, model, and parameters, to changes within Wikipedia between the fork on May 1, 2023, and January 1, 2025 for the same set of 56,835 articles, identified as deleted or heavily edited in the baseline comparison. If the narratives identified in the cross-platform comparison primarily capture Ruwiki's intervention rather than ordinary encyclopedic updating, the within-Wikipedia placebo should yield qualitatively different narratives. We therefore use it as a placebo benchmark for the main findings. We also apply the same methodology to changes within Ruwiki between the fork and January 1, 2025, allowing us to compare the narratives obtained from Ruwiki's own evolution with those from the baseline Ruwiki–Wikipedia comparison.

V.B Results: Three Pillars of Ruwiki's Censorship and Propaganda

We reproduce the output of the LLM narrative-extraction pipeline verbatim in the tables, without editing or reformulation by the authors, and provide our interpretation in the main text. Tables 1 and 2 report the narratives identified by GPT-4.1, with temperature set to 0, in the baseline comparison between Ruwiki and Wikipedia in January 2025. Table 1 presents the corpus-wide results. The top row provides the LLM's overall synthesis of the differences between the two corpora. Below it, the two columns report separately the narratives inferred from terms characteristic of deleted and inserted text. Table 2 presents corresponding assessments for a selected set of substantively important categories. We select categories that either account for a large share of heavily edited articles, such as history and politics and culture, media, and entertainment, or are especially overrepresented among heavily edited articles relative to their prevalence in Wikipedia, such as law and order and sex and gender. The category-specific results thus show how the corpus-wide narratives vary across the domains in which Ruwiki's intervention is quantitatively or substantively most salient.

We now interpret the machine-generated output. Taken together, the two tables reveal a coherent editorial intervention organized around three thematic pillars.

Pillar I: Aggression, war, and foreign policy. The first pillar concerns Russia's foreign policy and aggression, most prominently the war in Ukraine. At the corpus level, insertion narratives emphasize the formal incorporation of contested Ukrainian territories into the Russian Federation, their legal and administrative reorganization, and the official Russian justification of military operations. Deletion narratives are the mirror image: they remove explicit references to Russia's invasion and aggression, Ukrainian sovereignty and territorial integrity, Ukrainian administrative structures and local data, international sanctions and isolation, and Ukrainian historical and cultural context. The category-level results show that this pillar is especially prominent in *history and politics*, where additions legitimize Russian control over disputed territories and codify administrative changes, while deletions remove language that frames Russia as an aggressor, affirms Ukrainian sovereignty, or invokes international sanctions. The combined pattern normalizes Russian actions by shifting the encyclopedic record away from invasion, aggression, and Ukrainian sovereignty and toward territorial incorporation, legal-administrative reorganization, and the official Russian framing of military action.[21]

[21] This interpretation draws directly on the machine-generated narratives in Tables 1 and 2. The corpus-wide insertion narratives include "Integration of New Russian Federal Subjects," "Disputed Sovereignty and Ukrainian Contestation," "Legal and Administrative Reorganization," and "Military Operations and War Framing." The corresponding deletion narratives include "Removal of Invasion and Aggression Terminology," "Erasure of Ukrainian Sovereignty and Integrity References," "Removal

Pillar II: Domestic politics, opposition, and state authority. The second pillar concerns the suppression of domestic political dissent and the reinforcement of state authority. At the corpus level, the deletion narratives identify the removal of references to opposition figures, Western media and other critical sources, sanctions, and other dissenting perspectives. The category-level results show this pillar most clearly in *law and order*, where additions emphasize Russian legal sovereignty, traditional values, and a state-centered legal framing, while deletions remove references to human-rights violations, LGBTQIA+ rights, and independent or Western-aligned perspectives. The combined pattern shifts content away from opposition, independent scrutiny, and rights-based criticism and toward a state-centered legal and institutional framing.[22]

Pillar III: Culture, traditional values, and the West. The third pillar concerns culture, identity, and traditional values. At the corpus level, deletion narratives identify the removal of references to LGBTQIA+ identities, sexual orientation, and sexual content, while insertion narratives emphasize restrictive legislation, bans on so-called LGBTQIA+ "propaganda," and other negative framings of LGBTQIA+ issues. The category-level results show that this pillar is especially pronounced in *sex and gender*, where additions shift the treatment of gender and sexuality toward legal restriction and pathologization and frame LGBTQIA+ rights as foreign or threatening to traditional values, while deletions erase LGBTQIA+ visibility, rights, and representation in media, biographies, and legal contexts. It also appears in *law and order*, where additions emphasize traditional values, restrictions on LGBTQIA+ rights, and medicalized language, and in *culture, media, and entertainment*, where state-centric, nationalistic, and regulatory additions are accompanied by the removal of Western or liberal perspectives, critical reviews, and content concerning LGBTQIA+ rights. The combined pattern recasts gender and sexuality in state-centered, restrictive, and pathologizing terms, while presenting rights-based and international perspec-

---

of Ukrainian Administrative and Local Data," and "Omission of Sanctions and International Isolation." The category-level assessment for history and politics describes "a strong push to legitimize Russian control over disputed territories, codify administrative changes, and assert Russian legal and cultural norms," while deletions "systematically remove references to Russian aggression, Ukrainian sovereignty, international sanctions, and human rights issues." (For conciseness, we summarize the relevant category-specific machine-generated narratives in the accompanying footnotes rather than reproducing the full model output.)

[22] At the corpus level, the deletion narratives include "Suppression of Western and Opposition Sources": references to "Western media, opposition figures, and critical sources are deleted, reducing visibility of dissenting or critical perspectives." The category-level assessment for law and order describes additions "emphasizing Russian legal sovereignty, traditional values, and the pathologization of gender and sexual diversity," while deletions "systematically remove references to Russian aggression, human rights violations, LGBT rights, and independent or Western-aligned perspectives"; specific deletion narratives include "Suppression of Opposition and Independent Media" and "Censorship of Protest and Civil Society Activity."

tives as foreign or threatening to traditional values.[23]

Overall assessment. The three pillars are not orthogonal. They fit together as components of a single ideological project. The foreign-policy pillar legitimizes Russia's actions abroad by reframing invasion and annexation as lawful territorial incorporation. The domestic-politics pillar suppresses opposition and rights-based critique while reinforcing state authority. The cultural pillar presents Russia as a defender of traditional values against rights-based and Western liberal norms portrayed as foreign or threatening. Arguably, by promoting hierarchy, obedience, and respect for authority within the family and in gender relations, it reinforces acceptance of authority more broadly, including in the political sphere. The category-level analysis shows that this intervention is both coherent and targeted. Ruwiki applies a coherent ideological framework across the politically salient domains examined here, but the relative weight of the three pillars varies with subject matter: foreign-policy narratives dominate history and politics; state-authority and traditional-values narratives intersect in law and order; and cultural-conservative narratives are most pronounced in sex and gender, while also appearing in culture, media, and entertainment.

Stability We present evidence on the stability of the results in Online Appendix F. First, we show that the extracted narratives are highly stable as we vary the number of narratives requested by the prompt. Across all adjacent granularity levels of the full sweep from one to eight narratives per direction, and for both insertions and deletions, all 70 finer-level narratives (100%) preserve the hierarchy: they are classified as either still present or split; none is classified as new. Moreover, the dominant theme identified at the coarsest level remains dominant at every finer level. Online Appendix Figure A6 visualizes this structure: broad narratives in the inner rings subdivide into related, more specific narratives in the outer rings, with no new branches. Thus, the substantive conclusions do not depend on our preferred choice of seven narratives and do not arise from arbitrary reshuffling as the granularity of the prompt changes.

We then verify that the main narrative structure is stable across alternative LLMs and temperature settings. Together with the baseline GPT-4.1 temperature 0 specification, an independent replication of the

[23] The corpus-wide deletion narratives include "Censorship of LGBTQ+ and Sexual Content," while the insertion narratives include "LGBTQ+ Legislation and Propaganda Restrictions." The category-level assessment for sex and gender describes "legalistic, medicalized, and pathologizing framings of LGBT issues, with emphasis on restriction, criminalization, and the defense of traditional values," with LGBT rights portrayed as "foreign or dangerous," while deletions "systematically erase LGBT visibility, rights, and representation, especially in media, personal biographies, and legal contexts." The law-and-order assessment identifies addition narratives such as "Medicalization and Pathologization of Gender and Sexuality," "Highlighting Traditional Values and Cultural Distinctiveness," and "Linking LGBT Issues to Criminality and Social Danger." The culture, media, and entertainment assessment describes addition narratives focused on "Russian State and Legal Control in Media" and LGBT "propaganda" discourses, while deletion narratives target "Western and International Contexts," "Critical and Review-Based Discourse," and "LGBT Rights and Identities."

baseline specification, and twelve alternative model and parameter configurations identify substantively similar narratives, differing primarily in wording rather than substance. The narratives generated by the independent replication and the alternative models and temperatures are reported in Online Appendix Tables A2–A14. We conclude that the narratives presented above reflect stable patterns in the edited text, rather than artifacts of any single model or parameter choice.

V.C Placebo: Changes within Wikipedia

Could the patterns in Tables 1 and 2 reflect the ordinary evolution of Russian-language encyclopedic content over the same period, rather than Ruwiki's editorial intervention? For example, some cross-platform differences could arise as Wikipedia expanded its coverage of the war and other major events after the fork. To examine this possibility, we apply the identical pipeline to changes within the same Wikipedia articles between May 2023 and January 2025, leaving Ruwiki out of the analysis entirely.

The narratives identified by this within-Wikipedia placebo, reported in Online Appendix G (Online Appendix Table A15), are qualitatively different from those identified in the main analysis. Additions expand historical and military context, demographic and census information, cultural and religious heritage, coverage of sports and other achievements, and critical reception. Deletions primarily streamline articles by removing redundant or outdated administrative and demographic information, bureaucratic references, duplicative population data, and minor biographical and local details, and trim peripheral political and conflict-related content. Overall, the LLM characterizes Wikipedia's evolution as a combination of broad factual expansion and routine editorial maintenance.

Most importantly, the within-Wikipedia placebo does not exhibit the opposite-signed directional structure that would be required for ordinary Wikipedia evolution to generate the baseline results. To explain the terms characteristic of deletions from Ruwiki, Wikipedia would need to have systematically added language concerning Ukrainian sovereignty, Russian political opposition, rights-based criticism, and LGBTQIA+ identities and rights. To explain the terms characteristic of insertions into Ruwiki, Wikipedia would need to have systematically removed official territorial, state-centered legal, and traditional-values framings. The placebo reveals neither pattern. It does identify one limited exception: Wikipedia expanded its factual coverage of the war, including by adding references to Russia's invasion of Ukraine. As Ruwiki did not incorporate the same information, part of the divergence in war-related language reflects censorship by omission rather than the active deletion of pre-existing text. This secondary channel accounts for some war-related differences, but not for the broader combination of territorial legitimation, suppression of opposition and rights-based critique, and restrictive traditional-values narratives identified

in the baseline.

The placebo also addresses a methodological concern about LLM-based narrative extraction. If the pipeline mechanically produced a censorship-and-propaganda interpretation whenever it encountered Russian-language or politically salient terms, it would impose a similar structure on the within-Wikipedia changes, which include material on the war, sanctions, and other geopolitical events. Instead, it interprets these changes as broad factual updating and routine maintenance. The contrast therefore shows that the pipeline does not mechanically politicize ordinary Wikipedia evolution and supports the interpretation that the three-pillar structure captures Ruwiki's systematic editorial intervention.

Consistent with this interpretation, the within-Ruwiki comparison over time yields patterns very similar to those from the January 2025 cross-sectional comparison between Ruwiki and Wikipedia. This evidence indicates that active rewriting within Ruwiki was the principal driver of the narrative divergence between the two platforms over the twenty months following the fork, with selective non-updating providing an additional, more limited channel of censorship by omission (Online Appendix Table A16 versus Table 1).

VI How Are Narrative Changes Achieved?

The narrative pipeline in Section V characterizes Ruwiki's editorial intervention at the level of broad ideological themes. In this section, we examine the concrete editorial choices through which these narrative changes are produced. We first study which articles Ruwiki deletes altogether. We then turn to articles that remain in Ruwiki but are substantially rewritten.

VI.A What Articles Get Deleted?

We begin with the most extreme form of censorship: the wholesale removal of an article from the encyclopedia. Unlike heavy editing, which changes how a subject is presented while preserving the entry, deletion removes the topic entirely. As discussed in Section III, Ruwiki deleted 2,786 articles, only 0.14% of the 1.9 million main-namespace articles in the corpus. Because deletion is exceptional, the articles selected for removal are especially informative: they identify subjects that Ruwiki editors chose not to retain even in revised form, suggesting that reframing was deemed insufficient.

To characterize the substantive content of deletions, we rely primarily on the manual classification of deleted articles, which is designed to capture the apparent reason for removal. Figure 2 reports which topics are deleted and maps these manually coded topics to the broader LLM-based categories used for

the full corpus. We begin with the manual classification, since it provides the most direct evidence on which topics Ruwiki removes entirely.

Four substantive umbrella categories account for 91% of all deletions. *Culture* is by far the largest category, accounting for almost two-thirds of all deleted articles. Within it, the dominant fine-grained topic is pornography: 1,384 articles, nearly half of all deletions. These articles consist mostly of entries on adult-film performers, magazines, studios, and related organizations. The next largest cultural topic is LGBTQIA+ content, with 240 deleted articles on same-sex relationships, gender identity, LGBTQIA+ rights, and prominent LGBTQIA+ figures. Other cultural deletions include articles on drugs and substance use, sexuality and obscene language, feminism, suicide, and unconventional religious topics. Thus, the largest deletion category concerns cultural and moral content that Ruwiki appears to treat as incompatible with the regime's preferred values agenda, linking article deletions directly to the third pillar identified in Section V.B: the construction of Russia as a defender of traditional values.

War in Ukraine and contested geopolitics (labeled war/geopolitics) is the second-largest category, with 519 deleted articles. This group includes articles on the Russian invasion of Ukraine and specific battles within it, articles on war crimes and Russian military operations in Ukraine, and articles on geopolitical disputes in the former Soviet Union and Chechnya. These deletions are directly linked to the first pillar identified in Section V.B: the suppression of narratives of Russian aggression, wartime violence, and contested territorial claims.

Domestic politics in Russia forms a smaller but substantively important category, with 115 deleted articles. These include articles on the Russian political regime, human-rights organizations and activist movements, Aleksei Navalny and figures associated with the political opposition, restrictions on freedom of speech, radical organizations, and Russian propaganda institutions. These deletions correspond to the second pillar above: the suppression of domestic political dissent and criticism of state authority.

Controversial historical content accounts for approximately 120 deleted articles. These articles concern contested events in Russian, Ukrainian, and Soviet history, and many carry Wikipedia's "controversial history" label, which is typically applied to articles subject to active editorial dispute. This category connects historical memory to both foreign-policy and domestic-legitimacy narratives.

The remaining 9% of deletions fall into a residual category, consisting mostly of conventional encyclopedia-maintenance cases, such as page mergers, reorganizations, and articles flagged for citation problems. Ruwiki therefore performs some ordinary housekeeping, but such deletions are quantitatively small relative to the ideologically salient categories. We do not analyze them further.

Figure 2 also validates the LLM-based non-mutually exclusive topic classification against our more

detailed manual coding for the subset of articles where manual coding is feasible. The two classifications map closely onto one another. Deleted articles in the manual pornography topic flow overwhelmingly into the LLM categories *culture, entertainment, and media* and *people/person*, reflecting the large number of deleted biographies of adult-film actors. Similarly, articles in the manual war/geopolitics and domestic-politics umbrella groups map primarily into the LLM categories *history and politics* and *law and order*. The mapping shows that the broad LLM categories capture the same substantive structure as the manual classification, while also highlighting their limits: because the LLM classification is broad and overlapping, it identifies the substantive domain of an article but not necessarily the reason for deletion.[24]

Figure 3 reports the share of articles assigned to each LLM category in the full corpus, among deleted articles, and among heavily edited articles. Because categories are not mutually exclusive, these shares should be interpreted as category prevalence within each sample, not as a distribution that sums to 100%. Here, we focus on the deleted articles.

Deleted articles are concentrated in a narrow set of domains relative to the full corpus. The strongest overrepresentation is in *sex and gender*: this category accounts for only 1.4% of the full corpus but 60.4% of deleted articles, a difference of 59 percentage points and a 43-fold enrichment among deleted articles relative to the full corpus. *Culture, media, and entertainment* is also strongly overrepresented, accounting for 67.3% of deleted articles compared with 28.4% of the corpus; so are *people/person*, 63.1% compared with 35.0%; and *law and order*, 15.0% compared with 3.1%. By contrast, geography, sports, nature, environment, science, and technology are strongly underrepresented among deleted articles. As shown above, *history and politics* is substantively important, with 973 deleted articles, but it is not overrepresented relative to its large baseline prevalence in the corpus. Absolute differences in prevalence are statistically significant for all categories except *philosophy*.

Taken together, these descriptive patterns point to the same conclusion. Ruwiki rarely deletes articles outright. When it does, deletions concentrate in four ideologically salient domains: cultural and sexual content treated as morally subversive, especially pornography and LGBTQIA+ topics; articles documenting Russian military aggression, especially the war crimes committed by the Russian army during the invasion of Ukraine; articles documenting domestic political dissent and human-rights criticism, especially those related to Navalny's investigations; and articles whose treatment of Russian, Ukrainian, or Soviet history conflicts with the regime's preferred historical narrative. Deletion thus appears to be

[24]For example, *people/person* is a broad category, but among deleted articles its overlap with the manual classification is concentrated among pornographic actors and, to a lesser extent, politically sensitive figures. Similarly, *history and politics* is broad in the full corpus, but deleted articles within it are disproportionately about the war in Ukraine, Russian military actions, contested geopolitics, domestic opposition, and controversial historical memory.

reserved for topics that are difficult to reframe within an official narrative, or whose mere presence on the platform is inconsistent with that narrative.

VI.A.1 Article Characteristics That Predict Deletion

The manual classification provides the most direct interpretation of deleted topics, but it is necessarily judgment-based. We therefore complement it with a reproducible, text-based prediction exercise. The unit of observation is the article, and the dependent variable equals one if the article was deleted from Ruwiki and zero otherwise. The covariates are bigrams and trigrams extracted from the article's title and from its original Wikipedia category labels. Because deletions are rare and the feature space is high-dimensional, we estimate LASSO-regularized logistic regressions. To avoid basing the results on a single, potentially arbitrary sample of non-deleted control articles, we repeat the estimation across alternative control samples. In each of $B = 100$ iterations, we combine all 2,786 deleted articles with a random sample of 5,000 non-deleted articles and estimate a LASSO logit on the resulting sample. We retain for further analysis all n-grams with nonzero coefficients in at least 5% of the 100 iterations. Figure 4 and Online Appendix Tables A17 and A18 report only terms with coefficients of at least 0.05 in absolute value in at least 75% of the iterations, ensuring that the reported predictors are stable across alternative control samples. Online Appendix H provides further details on the empirical methodology and the cross-validation procedure used to select the penalty parameter λ.

Figure 4 reports the most consistently selected bootstrap-stable n-grams translated into English, together with their coefficients. The results closely match the manual classification. The strongest predictor of deletion is *pornography article*, followed by *LGBTQIA+* article. These terms correspond to the largest part of the manual culture category and capture the most distinctive deletion component of the cultural pillar: the removal of sexual and LGBT-related content treated as morally unacceptable. A second group of predictors, including *Russian invasion*, *Russian war*, *Invasion of Ukraine*, *Battle territory*, *War crime*, *During the invasion*, *Russian-led invasion*, and *Russo-Ukrainian war*, identifies articles documenting the invasion of Ukraine, Russian military operations, and associated atrocities. Finally, terms such as *Russian Federation* and *caution editing*, Wikipedia's contributor-risk for politically sensitive content, point to the domestic-politics and geopolitics clusters.[25] Thus, the machine-selected predictors reproduce the same substantive deletion domains identified by the manual coding: culture and sexuality, war and geopolitics, domestic politics, and contested history. The results are robust to using categories only, titles only, titles

[25] Among the deleted articles, 88 carry a contributor-risk notice. Of these, 65 fall within the war/geopolitics manually-coded umbrella group, 18 within domestic politics, four within culture, and one within "other" umbrella group.

plus categories, with and without unigrams, as documented in Online Appendix I. In contrast to Ruwiki, Wikipedia did not delete any main namespace articles.

Overall, deletions reproduce the three-pillar structure identified in Section V. At the same time, they represent only a small share of Ruwiki's editorial intervention. Most editorial work occurs inside retained articles, through the selective insertion and deletion of phrases, sentences, and paragraphs. We turn to that machinery next.

VI.B What Articles Are Heavily Edited?

We now turn to articles that remain on the platform but are substantially rewritten. This is the main form of editorial intervention in Ruwiki: heavily edited articles outnumber deleted articles by a factor of 19.4 (54,049 versus 2,786). Heavy editing is therefore the central mechanism through which Ruwiki changes the encyclopedic record. It allows editors to preserve the appearance of coverage while altering the framing, terminology, and substantive interpretation of politically or morally sensitive topics. Arguably, this strategy reduces the risk that users will turn to independent platforms when they cannot find information on a particular topic.

Illustrative examples. The systematic changes are best understood through concrete examples. Online Appendix J reproduces translated excerpts from six articles for illustration purposes. The Ruwiki version of *LGBT Rights in Russia*, corresponding to the cultural pillar, adds that it is "the West" that treats LGBT issues as a problem, that "propaganda of same-sex relations in Russia is prohibited by law," and that "most Russian lawyers and human rights activists believe" that LGBT representatives "want special rights for themselves." At the same time, it deletes text stating that the "invisibility and infringement" of LGBT representatives "in Russia are perceived as the norm" and "are associated with aggression, intolerance and political repression against LGBT in Russia" (Online Appendix Figure A7).

The Ruwiki version of the article on the *Katyn massacre* illustrates the foreign-policy pillar in historical perspective. It adds that "to this day, the question of who was guilty of this crime remains the subject of scholarly and public debate," while deleting the statement that the executions "were carried out by the NKVD of the USSR in accordance with the resolution of the Politburo of the Communist Party Central Committee" (Online Appendix Figure A8).

The next three examples, *Vladimir Putin*, *Aleksei Navalny*, and *Human Rights in Russia*, illustrate the domestic-politics pillar (Online Appendix Figures A9–A11). These articles are not merely adjusted at the margin; they are substantially rewritten. In the *Vladimir Putin* article, the volume of inserted text

is roughly comparable to the volume of deleted text. In *Aleksei Navalny*, most of the original text is removed, while the additions characterize Navalny as a video blogger, an extremist, and a member of the "so-called Russian non-parliamentary opposition." In *Human Rights in Russia*, deletion is the dominant operation: most of the text is removed, producing a substantially less critical account of the human-rights situation.

A particularly telling example of intervention in domains of culture and history is the article on George Orwell's novel *1984*, in which Ruwiki severs the dystopia from its Soviet referent. It deletes the statement that Orwell modeled the totalitarian state "after the Stalinist USSR and Nazi Germany" and instead presents the model as contested between the Stalin-era USSR and 1940s Britain. It also adds the claim that "Big Brother fits the image of a Western ruler" and deletes a reference to a film about Gareth Jones, the journalist who exposed the Stalin-engineered famine in Soviet Ukraine known as the Holodomor (Online Appendix Figure A12).

VI.B.1 Which Article Categories Are Heavily Edited?

We use the LLM category prevalence reported in Figure 3 to characterize which retained articles are heavily edited. Heavily edited articles display a broader pattern than deleted articles. In absolute terms, they are most concentrated in *history and politics* (50.2%), *people/person* (38.8%), *culture, media and entertainment* (36.9%), and *geography* (24.8%), but these are also relatively large categories in the full corpus. Relative to the corpus, the largest percentage-point overrepresentations are in *culture, media, and entertainment* (+8.5 percentage points (pp)), *history and politics* (+6.9 pp), *people/person* (+3.8 pp), *technology* (+3.3 pp), and *law and order* (+3.2 pp).

Smaller categories are especially revealing in relative terms: *sex and gender* is 2.7 times as prevalent among heavily edited articles as in the full corpus, *psychology and psychiatry* 2.1 times, and *law and order* 2.0 times. Thus, heavy editing combines broad intervention in large categories, especially history and politics and culture, with disproportionate attention to smaller but ideologically salient domains. Even categories that are strongly underrepresented among heavily edited articles remain substantively important in absolute terms: roughly 13,383 geography articles, 4,002 sports articles, and 3,307 nature articles are heavily edited. Absolute differences in prevalence are statistically significant for all categories.[26]

[26] Appendix Figure A2 shows that, for the 522,135 articles with any edit, the prevalence of LLM categories generally lies between the corresponding prevalence among heavily edited articles and that in the full corpus. The only exceptions are *people/person* and *sports*, which are more prevalent among all edited articles than among either heavily edited articles or the full corpus. Appendix Figure A3 zooms into articles that carry a contributor-risk notice. It shows that both heavily edited and all edited such articles are much more concentrated among *history and politics*, *geography*, *people/person*, *law and order*, and *culture, media, and entertainment*.

VI.B.2 Textual Features of Insertions and Deletions in Heavily Edited Articles

The category analysis tells us *which* articles attract editorial attention. To understand *how* Ruwiki rewrites these articles, we move to the level of individual text segments. A segment is a contiguous block of text present on one platform but not on the other. We classify a segment as an *insertion* if it is present in Ruwiki but absent from Wikipedia, and as a *deletion* if it is present in Wikipedia but absent from Ruwiki. Pooling segments across heavily edited articles yields 838,219 text segments from 49,926 articles whose cleaned text passes our preprocessing pipeline.[27]

We then estimate a LASSO logit in which the dependent variable equals one for insertions and zero for deletions. The covariates are binary indicators for lemmatized bigrams and trigrams appearing in each segment. The specification includes penalized article indicators, which account for differences in topics and editing intensity across articles while allowing these article-specific effects to be regularized. The L1 penalty is applied jointly to the n-gram coefficients and the article indicators, and the penalty parameter is selected by 10-fold cross-validation. The selected n-grams therefore identify phrases associated with the direction of editorial change after controlling, in a regularized manner, for article-level heterogeneity. Online Appendix K.1 reports the full specification.

As above, we use the within-Wikipedia evolution between the fork in May 2023 and January 2025 as a placebo to distinguish active Ruwiki editing from ordinary temporal change. Online Appendix K.2 applies our conceptual framework to the baseline and placebo comparisons. For the same set of heavily edited articles, it separates the selected LASSO terms into four groups with distinct interpretations. First, terms selected only in the contemporaneous Ruwiki–Wikipedia comparison are the most direct markers of active divergence between the two platforms. Second, terms selected only in the within-Wikipedia placebo characterize ordinary evolution of Wikipedia content. Third, terms selected in both comparisons with opposite signs are consistent with Ruwiki staleness or censorship by omission: Wikipedia changed after the fork, whereas Ruwiki retained the earlier wording or failed to incorporate the new material. Finally, terms selected in both comparisons with the same sign indicate that both platforms updated in the same direction, but with different intensity or emphasis. We therefore report four groups of terms: baseline-only terms, placebo-only terms, and terms appearing in both comparisons with either opposite or identical signs.

We begin with terms selected by the LASSO in the baseline Ruwiki–Wikipedia comparison but not

[27] The article count is below the total of 54,049 heavily edited articles because the document–term matrix used by the LASSO requires an article to contain at least one bigram or trigram surviving our part-of-speech and frequency filters. The 4,123 excluded articles are short or consist primarily of proper nouns or numeric content. See Online Appendix A.3.

in the within-Wikipedia placebo. These provide the clearest markers of active, Ruwiki-specific editorial intervention. To focus on the largest associations, we report up to ten terms in each direction with an estimated LASSO coefficient above 1.4 in absolute value. Panel A of Figure 5 reports the coefficients for the full sample of heavily edited articles. The phrases most strongly predictive of deletion are dominated by the war in Ukraine and the status of contested territories.[28] The leading deletion predictors also include *Gay Lesbian*, associated with cultural censorship; *Novaya Gazeta*, Russia's preeminent independent newspaper, whose editor-in-chief received the Nobel Peace Prize; and *Alexey Navalny*, the Russian opposition leader, associated with domestic-politics censorship.

The phrases most strongly predictive of insertion are likewise primarily geopolitical. They replace the language of invasion, Ukrainian sovereignty, and contested territorial status with official, legal, and administrative terminology.[29] The appearance of *Kosovo crisis* among the leading insertion predictors suggests an additional strategy of relativization, in which Ruwiki frames Russia's actions in Ukraine by invoking other contested international cases. Another characteristic insertion predictor across several sensitive domains is *Banned in the Russian Federation*, a broad legalistic formula used to designate organizations, materials, and activities as unlawful. In the cultural domain, the leading inserted term is *LGBT propaganda*, a regime-sanctioned legalistic label. Taken together, these terms show that references to Russian aggression and Ukrainian sovereignty are removed and replaced with regime-sanctioned terminology concerning the war, territorial administration, and prohibited organizations. These are precisely the kinds of words and phrases that Russian officials have demanded that Wikipedia change (Fadeyev, 2023). The same pattern extends beyond the war: LGBTQIA+ content is reframed through the language of so-called "LGBT propaganda," while references to opposition figures and independent media are removed.

Panels B–D re-estimate the same LASSO logit separately within three heavily targeted LLM-defined categories—history and politics, law and order, and sex and gender—in which the distinctive pattern of Ruwiki intervention is particularly visible. In *history and politics* (Panel B), the pattern closely mirrors the full-corpus result. In *law and order* (Panel C), the same military and territorial patterns remain visible but are joined by additional themes. In the regulation of gender and sexuality, the leading deletion predictors include terms such as *intersex rights*, whereas the insertion predictors include terms such as *intersex*

[28] They include several phrases related to Russia's invasion of Ukraine, such as *Russian invasion*, *invasion of Ukraine*, *Russo-Ukrainian war*, and *Luhansk oblast*, the Ukrainian administrative designation for an occupied region.

[29] The associated terms include *special military operation*, the Kremlin's official designation for the war in Ukraine; *military actions Ukraine*, used in place of invasion; *region territory of Ukraine*, reflecting Russian administrative classifications applied to annexed territories; and *continue consider territory*.

registration. In domestic politics, the inserted text includes *the list of foreign agents*, referring to an official designation used to restrict members of the opposition and civil society. The prominence of *Kosovo crisis* and *Kosovo war*, together with *many countries of the world*, reinforces the interpretation that the regime pursues a strategy of relativization. Thus, in law-and-order articles, Ruwiki combines the reframing of military and territorial issues with the language of legal prohibition, comparative justification, and the regulation of gender and sexuality. In *sex and gender* (Panel D), the cultural pillar appears in its clearest form. Deleted text disproportionately contains the language of rights, recognition, and protected identities, whereas inserted text disproportionately contains the language of pathology, regulation, and prohibition. For example, terms such as *recognition of same-sex relationships* and *intersex rights* appear among the deletion predictors, whereas *LGBT propaganda* and *intersex registration* appear among the insertion predictors. This is the category in which Ruwiki's traditional-values narrative is most explicit. Online Appendix Figure A13 reports the results for other large LLM-defined categories—culture, media, and entertainment; people; geography; and technology. Across these categories, the pattern is broadly similar.

Within-Wikipedia placebo, phrase-level editing. The within-Wikipedia placebo for the same set of heavily edited articles produces a qualitatively different pattern. Online Appendix Figures A14 and A15 report terms selected in the within-Wikipedia comparison but not in the contemporaneous Ruwiki–Wikipedia comparison. Unlike the baseline terms, these phrases do not cohere around a common political or ideological theme. Instead, they largely reflect content that emerged during the period, together with routine updating and editorial cleanup. Moreover, in categories with relatively little event-driven updating—such as law and order, sex and gender, and technology—fewer than ten terms often meet the threshold of an absolute editing LASSO coefficient above 1.4. By contrast, the baseline comparison generally produces a full set of ten terms in each direction.

The third group consists of terms selected in both the baseline Ruwiki–Wikipedia and within-Wikipedia comparisons with opposite signs. These terms also mostly reflect routine Wikipedia updates that Ruwiki did not incorporate. However, two terms appearing both in the full sample of heavily edited articles and in the history-and-politics category point more directly to censorship by omission: Wikipedia added text related to the *annexation of Crimea* and *territory transfers*, whereas Ruwiki did not incorporate these additions. See Online Appendix Figures A16 and A17. Only one term appears in both comparisons with the same sign. It reflects updating of Ukraine's administrative classification system.[30]

[30] Ukraine's outdated KOATUU administrative classification was replaced by the newer KATOTTH system. Both platforms incorporated this change, but Ruwiki did so more extensively.

Consistent with our interpretation of the LASSO editing results, a complementary held-out classification exercise shows that the selected vocabulary separates Ruwiki from contemporaneous Wikipedia substantially more sharply than it separates Wikipedia from its fork-date version (see Online Appendix Section L.2 and Figure A18).[31]

Overall, the LASSO results provide a phrase-level account of the narrative patterns documented above. Ruwiki editors do not simply delete content they deem politically or morally sensitive. They replace it with official, legalistic, and bureaucratic terminology that reclassifies the same events, territories, organizations, and identities within the regime's preferred vocabulary.

VI.B.3 Readability Costs of Ruwiki Editing

We next ask whether and how censorship and propaganda affect the quality and readability of the resulting text. The effect is ambiguous ex ante. Because Ruwiki is edited by professionals rather than volunteers, its editors might improve the clarity and readability of articles. Conversely, replacing Wikipedia text with the official, legalistic, and bureaucratic formulations could produce denser and less accessible prose. We examine five outcomes: text length, measured by the number of Russian words; sentence length, measured by words per sentence; lexical complexity, measured by average word length; readability, measured by the Russian-language adaptation of the Flesch Reading Ease score (Flesch, 1948; Oborneva, 2006); and lexical diversity, measured by MATTR with a 50-word moving window (Covington and McFall, 2010).

We calculate these measures using the complete, unabridged Wikipedia and Ruwiki versions of each heavily edited article. We also examine whether the text-quality differential between the two sources is larger for topics that are particularly sensitive to the Russian regime. For this purpose, we use a marker supplied by Wikipedia itself. A contributor-risk notice warns editors that contributing to an article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. In the pooled sample of Wikipedia and Ruwiki versions, we regress each text-quality outcome on indicators for Ruwiki and the contributor-risk notice and their interaction. In specifications with article fixed effects, all constant differences across articles, including the main effect of the presence of the contributor-risk notice, are absorbed.

Table 3 presents the results. Panel A reports specifications without article fixed effects, while Panel B

[31] As a robustness check, we replace the LASSO logit with a non-parametric measure of association: signed Cramér's V, equivalently the scaled ϕ coefficient, computed from the 2×2 contingency table of n-gram presence by edit type. The two methods produce closely matching results. Online Appendix M presents the terms that are most characteristic of the difference between Ruwiki and Wikipedia, our baseline comparison (Appendix Table A19) and the terms that are most characteristic of the within-Wikipedia, our placebo comparison (Appendix Table A20). In both tables we omit the terms that appear in both comparisons due to ambiguity of their interpretation (see Appendix Sections E and K.2).

compares the Ruwiki and Wikipedia versions of the same article as it controls for article fixed effects. For heavily edited articles without a contributor-risk notice, Ruwiki editing has only limited effects on text quality; its main effect is to remove content. The Ruwiki version contains 183 fewer words, a reduction of about 23% relative to the Wikipedia mean. Because article length is highly dispersed, however, this difference amounts to only 0.14 of the Wikipedia standard deviation (SD). The remaining differences are small in magnitude, although statistically significant. With the exception of slightly shorter sentences, they generally point toward lower text quality: sentence length falls by 0.3 words, or 0.04 SD; average word length rises by 0.07 SD; the Flesch score falls by 0.03 SD; and MATTR-50 falls by 0.02 SD. Thus, for heavily edited articles outside the most sensitive set of topics, Ruwiki substantially prunes content while only slightly worsening the readability and lexical properties of the remaining text.

The results are strikingly different for heavily edited articles carrying a contributor-risk notice. The interaction coefficients, which capture the additional effect of Ruwiki editing on these articles relative to those without a notice, are large and precisely estimated. Combining the Ruwiki coefficient with the interaction, contributor-risk articles contain approximately 312 fewer words on Ruwiki than on Wikipedia, a reduction equal to about 40% of the overall Wikipedia mean. More importantly, the retained text becomes substantially harder to read. Sentences are 2.02 words longer, equivalent to 0.29 SD; average word length rises by 0.19 characters, or 0.37 SD; the Flesch score falls by 9.17 points, or 0.53 SD; and MATTR-50, multiplied by 100, falls by 5.3 points, equivalent to 0.93 SD. These effects are therefore not only statistically precise but also substantively large. The deterioration in text quality is large among articles for which contributing to Wikipedia may expose editors to legal or personal risk in Russia or Belarus.

The results are robust to the inclusion of article fixed effects: the estimates are very similar in Panels A and B. They are not merely a mechanical consequence of the shorter length of Ruwiki articles. Panel C controls for the log number of words and yields similar estimates. Because article length is itself an outcome of Ruwiki editing, this specification conditions on an endogenous post-treatment variable and should therefore be interpreted with caution. Nevertheless, it provides useful evidence that intensive Ruwiki editing of sensitive topics produces denser, more complex, less varied, and markedly less readable text even conditional on article length.

A within-Wikipedia placebo indicates that these cross-platform differences reflect deterioration following Ruwiki intervention rather than improvements in Wikipedia after the fork. Comparing the same heavily edited articles in the May 2023 and January 2025 Wikipedia snapshots, we find changes in text quality that are very small in magnitude and mixed in direction, including among contributor-risk articles, and none remotely comparable to the differences observed in Ruwiki–Wikipedia comparision (as

reported in Online Appendix Table A22 and discussed in detail in Online Appendix Section N).

We also replicate the cross-platform analysis for edited articles that do not meet our threshold for heavy editing. The cross-platform differences in readability are negligible in this subsample, providing no evidence that professional Ruwiki editing improves text quality even at the margin and confirming that the substantial deterioration documented above is concentrated among articles subject to extensive intervention (Online Appendix Section N).

Overall, readability deteriorates substantially when Ruwiki intervenes extensively in articles that Wikipedia identifies as especially sensitive, pointing to the substantial readability cost of information control. This finding reinforces the phrase-level evidence presented above. Censorship and propaganda thus affect not only which facts and narratives are presented, but also how accessible the text is to readers.

VII Official Russian History Textbooks as an External Benchmark

We next provide external validation that Ruwiki's edits move historical content toward the Russian state's preferred narrative. We compare Wikipedia and Ruwiki history articles with Russia's official Grade-11 history textbooks covering the same topics. The textbooks cover Russian and world history from the end of World War II to the present (Medinsky and Torkunov, 2023; Medinsky and Chubar'yan, 2023).

We build on the narrative-comparison procedure developed by Waight et al. (2025), adapting it to identify contradictions rather than overlapping claims. This distinction is important in the context of history: our objective is to identify where accounts of the same historical events diverge, rather than where they share factual content.

To compare the encyclopedias with the textbooks on the same subjects, we OCR-digitize the two official Grade-11 history textbooks, divide them into topical subchapters, and match each heavily edited Wikipedia history article to the most similar textbook subchapter using embedding similarity. We retain the 100 highest-similarity matches. Manual verification confirms that these 100 article–subchapter pairs cover the same topics; beyond the top 100, the matches become increasingly spurious. The resulting dataset contains 100 matched triples, each consisting of a textbook subchapter and the Wikipedia and Ruwiki versions of the corresponding article.

For each text, following Waight et al. (2025), we use an LLM to summarize it, translate the summary, and extract substantive causal, normative, and conceptual claims from the translated summary. The model extracts, on average, 11.8 claims per Wikipedia article, 11.4 per Ruwiki article, and 21.3 per textbook subchapter. We then ask the LLM to identify claims in each of the 200 encyclopedia versions that directly contradict claims in the matched textbook subchapter. We identify 33 direct contradictions in 25 of the

200 encyclopedia versions: 25 contradictions across 19 Wikipedia versions and 8 contradictions across 6 Ruwiki versions.

We test whether Wikipedia and Ruwiki differ significantly in the incidence and number of contradictions using paired regression specifications and exact methods appropriate for small samples. Online Appendix O provides details on the data-construction pipeline, illustrative examples, and empirical strategy.

In the regression analysis, the dependent variable is either an indicator for whether an encyclopedia version contains at least one contradiction with the matched textbook subchapter, estimated using a linear probability model, or the number of such contradictions, estimated using PPML. The key explanatory variable is an indicator equal to one for Wikipedia and zero for Ruwiki. We progressively add controls for the numbers of substantive claims extracted from the encyclopedia version and the corresponding textbook subchapter, followed by article-pair fixed effects. Because the Wikipedia and Ruwiki versions in each pair are evaluated against the same textbook subchapter, the fixed effects absorb the number of claims in that subchapter. Our preferred specification includes article-pair fixed effects, so the coefficient on the Wikipedia indicator compares the Wikipedia and Ruwiki versions of the same article against the same textbook benchmark. Standard errors are clustered by article pair.

Table 4 reports the regression results. Wikipedia articles contradict the official textbook narrative significantly more often than their Ruwiki counterparts. In the specification with all controls, the Wikipedia version of the average matched article is 13 percentage points more likely to contain at least one contradiction ($p = 0.001$) and contains approximately three times as many contradictions as the corresponding Ruwiki version ($p = 0.007$).

Given the sparsity of contradictions, we complement the regression analysis with exact paired tests and within-pair permutation tests, reported in Online Appendix Table A25. For the any-contradiction outcome, 15 matched pairs contain at least one contradiction in Wikipedia but none in Ruwiki, whereas only two pairs exhibit the reverse pattern. An exact two-sided binomial test rejects symmetry among these discordant pairs ($p = 0.002$). A within-pair permutation test that randomly swaps the Wikipedia and Ruwiki labels likewise rejects equality for this outcome ($p = 0.002$). For the contradiction-count outcome, the exact paired sign test and the within-pair permutation test also reject the null of no systematic within-pair difference, with p-values of 0.008 and 0.001, respectively.

The matching design is conservative with respect to our hypothesis. Article–subchapter matches are selected using the Wikipedia version rather than the Ruwiki version, so the procedure mechanically favors Wikipedia in its topical similarity to the textbook benchmark. Any resulting bias should therefore

work against finding that Ruwiki is closer to the textbook. We verify this by constructing an alternative sample in which the top 100 matches are selected using the Ruwiki rather than the Wikipedia version and reestimating the analysis. The resulting point estimates are consistently larger in magnitude (Online Appendix Table A26).

A within-Wikipedia placebo confirms that the results are not driven by Wikipedia moving away from the textbook narratives after the fork. Using the same 100 matched articles as in the baseline analysis, we compare contradiction incidence and counts between the May 2023 and January 2025 Wikipedia versions. The differences are both statistically insignificant and substantively small (Online Appendix Table A27).

Taken together, these results show that Ruwiki versions are substantially less likely to contradict the official textbook narrative, providing external validation that Ruwiki's revisions move historical content toward the Russian state's preferred narrative.

VIII Conclusion

Understanding authoritarian information control requires identifying the content autocratic regimes suppress, the narratives they promote, and the ways in which these interventions fit together into a coherent worldview. This paper examines how an authoritarian regime rewrites a comprehensive body of public knowledge. The creation of Ruwiki, a regime-aligned fork of Russian-language Wikipedia, allows us to compare two encyclopedias that began with identical content but subsequently evolved under sharply different institutional control. This setting makes it possible to observe censorship and propaganda jointly, at scale, and across the full range of encyclopedic knowledge.

Two complementary findings support our characterization of Ruwiki as a regime-sponsored project. Its editing patterns are difficult to reconcile with decentralized volunteer production and instead point to an organized, schedule-driven editorial operation. An external comparison with official high-school history textbooks further shows that Ruwiki's revisions move accounts of modern history toward state-sanctioned narratives.

Our main findings characterize the content and structure of Ruwiki's censorship and propaganda. It is organized around three main priorities: Russia's image vis-à-vis the rest of the world, domestic politics and state authority, and culture and traditional values. Across these areas, Ruwiki suppresses accounts of invasion, Ukrainian sovereignty, domestic opposition, human-rights violations, and liberal social norms, while introducing official, legalistic, and bureaucratic language that legitimizes territorial expansion, reinforces state authority, and presents Russia as a defender of traditional values.

Four broader lessons emerge. First, culture is not peripheral to authoritarian information control.

It is one of its central domains. Narratives about morality, gender, family, and traditional values are likely both to reinforce respect for political authority within Russia and to present Russia as the defender of a traditional moral order in opposition to the West. The prominence of culture therefore shows that propaganda seeks to shape not only views about specific policies or events, but also the moral framework through which political authority is understood.

Second, outright deletion is not the regime's principal instrument of information control. Ruwiki deletes relatively few articles, and those deletions are concentrated in topics that are especially difficult to reframe, including pornography, war crimes in Ukraine, corruption investigations, and contested historical memory. Most intervention instead takes place within retained articles. This strategy preserves the appearance of comprehensive coverage while changing the interpretive frame through which sensitive subjects are presented. It is also consistent with a preference for supplying a regime-compatible answer rather than leaving an unanswered query that might lead users elsewhere.

Third, the intervention is encyclopedic rather than confined to conventionally political content. Ruwiki advances its priorities concerning foreign policy, domestic authority, and traditional values not only in articles explicitly about politics, history, culture, or geography, but also in domains including law, gender, psychology, philosophy, science, technology, nature, and sports. The intervention therefore does not target only a limited set of politically sensitive claims; it seeks to align a much broader system of knowledge with the regime's worldview.

Fourth, ideological control comes at a cost to textual quality. On politically sensitive topics, Ruwiki replaces standard encyclopedic language with official terminology, legalistic formulations, bureaucratic classifications, and euphemisms. These substitutions make edited articles less readable.

Taken together, our findings suggest that censorship and propaganda are best understood as a single editorial project. Ruwiki does not simply remove information that is inconvenient to the regime. It constructs a coherent alternative representation of Russia's past and present, linking foreign policy, domestic authority, social hierarchy, and moral order. The result is not an encyclopedia with conspicuous gaps, but a broad, internally consistent, and systematically distorted knowledge system in which familiar topics remain available while their meanings are altered.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *The Quarterly Journal of Economics*, 130(4): 1885–1939.

Aristotle. ca. 350 BCE. *The Politics.*

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2024. "Relatio: Text semantics capture political and economic narratives." *Political Analysis*, 32(1): 115–132.

Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer. 2026. "GPT as a Measurement Tool." National Bureau of Economic Research Working Paper 34834.

BBC Russian. 2023. "Rewriting History: New Russian History Textbooks." *https://bbcrussian.substack.com/p/rewriting-history-new-russian-history-textbooks*, Accessed: 2025-01-03.

Besley, Timothy, and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review*, 96(3): 720–736.

Blouin, Arthur, and Sharun W. Mukand. 2019. "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda." *Journal of Political Economy*, 127(3): 1008–1062.

Boleslavsky, Raphael, Mehdi Shadmehr, and Konstantin Sonin. 2021. "Media Freedom in the Shadow of a Coup." *Journal of the European Economic Association*, 19(3): 1782–1815.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review*, 109(6): 2294–2332.

Covington, Michael A., and Joe D. McFall. 2010. "Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR)." *Journal of Quantitative Linguistics*, 17(2): 94–100.

DeepSeek-AI, Aixin Liu, Bei Feng, Bingxuan Xue, et al. 2024. "DeepSeek-V3 Technical Report." *arXiv preprint arXiv:2412.19437*.

DellaVigna, Stefano, and Eliana La Ferrara. 2015. "Economic and Social Impacts of the Media." In *Handbook of Media Economics*. Vol. 1, 723–768. Elsevier.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2(1): 643–669.

Deutsche Welle. 2023. "Russia-Ukraine War: How It's Being Taught in School History Books." *https://www.dw.com/en/russia-ukraine-war-history-school-books/a-66498211*, Accessed: 2025-01-03.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The Political Legacy of Entertainment TV." *American Economic Review*, 109(7): 2497–2530.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *The Review of Economic Studies*, 80(4): 1422–1458.

Egorov, Georgy, and Konstantin Sonin. 2024. "The Political Economics of Non-democracy." *Journal of Economic Literature*, 62(2): 594–636.

Ellul, Jacques. 1965. *Propaganda: The Formation of Men's Attitudes.* New York:Alfred A. Knopf. Translated from the French by Konrad Kellen and Jean Lerner.

Enikolopov, Ruben, and Maria Petrova. 2015. "Media Capture." In *Handbook of Media Economics*. Vol. 1, 687–700. Elsevier.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–3285.

EUvsDisinfo. 2026. "The Wikipedia Wars." *StopFake*, https://www.stopfake.org/en/the-wikipedia-wars/, Accessed May 21, 2026.

Fadeyev, Valery. 2023. "Statement on Wikipedia and Ukraine coverage." Chairman of the Presidential Council for Civil Society and Human Rights. As reported in *Reuters* and *Interfax*, April 2023.

Flesch, Rudolf. 1948. "A New Readability Yardstick." *Journal of Applied Psychology*, 32(3): 221–233.

Foreign Policy. 2023. "Russia's New Schoolbook Aims to Indoctrinate Children With Hatred for the West." *https://foreignpolicy.com/2023/09/03/russia-schoolbook-schools-ukraine-west-indoctrination-history-falsification-war-hatred-putin/*, Accessed: 2025-01-03.

Gehlbach, Scott, and Konstantin Sonin. 2014. "Government control of the media." *Journal of Public Economics*, 118: 163–171.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.

Greenstein, Shane, and Feng Zhu. 2012. "Is Wikipedia Biased?" *American Economic Review*, 102(3): 343–48.

Guriev, Sergei, and Daniel Treisman. 2019. "Informational Autocrats." *Journal of Economic Perspectives*, 33(4): 100–127.

Guriev, Sergei, and Daniel Treisman. 2020. "A Theory of Informational Autocracy." *Journal of Public Economics*, 186: 104158.

Guriev, Sergei, and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century.* Princeton University Press.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. "spaCy: Industrial-strength Natural Language Processing in Python."

Interfax. 2022. "Поисковики будут помечать "Википедию" как нарушителя законодательства РФ." *https://www.interfax.ru/russia/853035*.

Khimshiashvili, Polina, and Petr Kanaev. 2024. "Правительство определилось с будущим Большой российской энциклопедии." *RBC*, https://www.rbc.ru/politics/04/10/2024/6620c8f79a7947fd050a0da0, Accessed May 21, 2026.

Khinshtein, Alexander. 2023. "Statement on Wikipedia and domestic alternatives." Chair of the State Duma Committee on Information Policy. Posted on Telegram, June 28, 2023. As reported by *Bloomberg* and *Reuters*.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review*, 107(2): 326–343.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science*, 345(6199): 1–10.

Klemperer, Victor. 2000. *The Language of the Third Reich: LTI—Lingua Tertii Imperii: A Philologist's Notebook.* London:The Athlone Press. Translated by Martin Brady.

Kommersant. 2026. ""Ruviki" zapretila redaktirovat' stat'i pol'zovatelyam bez verifikatsii." *Kommersant*, Accessed May 27, 2026.

Korobov, Mikhail. 2015. "Morphological Analyzer and Generator for Russian and Ukrainian Languages." In *Analysis of Images, Social Networks and Texts*. Vol. 542 of *Communications in Computer and Information Science*, , ed. Mikhail Yu. Khachay, Natalia Konstantinova, Alexander Panchenko, Dmitry I. Ignatov and Valeri G. Labunets, 320–332. Springer International Publishing.

Kurek, Laura, Ceren Budak, and Eric Gilbert. 2025. "Wikipedia in Wartime: Experiences of Wikipedians Maintaining Articles About the Russia–Ukraine War." *Proceedings of the ACM on Human-Computer Interaction*, 9(2).

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science*, 58(2): 402–414.

Lu, Yingdan, and Jennifer Pan. 2021. "Capturing Clicks: How the Chinese Government Uses Clickbait to Compete for Visibility." *Political Communication*, 38(1-2): 23–54.

Mattingly, Daniel C., and Elaine Yao. 2022. "How Soft Propaganda Persuades." *Comparative Political Studies*, 55(9): 1569–1594.

Medinsky, Vladimir R., and Alexander O. Chubar'yan. 2023. *Vseobshchaya istoriya. 1945 god – nachalo XXI veka: 11 klass: bazovyi uroven': uchebnik [World History, 1945 to the Beginning of the 21st Century: Grade 11].* Moscow:Prosveshchenie.

Medinsky, Vladimir R., and Anatoly V. Torkunov. 2023. *Istoriya Rossii. 1945 god – nachalo XXI veka: 11 klass: bazovyi uroven': uchebnik [History of Russia, 1945 to the Beginning of the 21st Century: Grade 11].* Moscow:Prosveshchenie.

Miller, Carl, Melanie Smith, Oliver Marsh, Kata Balint, Chris Inskip, and Francesca Visser. 2022. "Information Warfare and Wikipedia." Institute for Strategic Dialogue and CASM Technology.

Mill, John Stuart. 1869. "The Subjection of Women." In *Essays on Equality, Law, and Education*. Vol. 21 of *The Collected Works of John Stuart Mill*, , ed. John M. Robson, 259–340. Toronto:University of Toronto Press. Reprinted in the 1984 edition.

Novaya Gazeta Europe. 2025. "Ruviki vedut k Kovalchuku i Banku Rossiya." Available at: https://novayagazeta.eu/articles/2025/05/20/ruviki-vedut-k-kovalchuku-i-banku-rossiia.

Oborneva, Irina V. 2006. "Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov." PhD diss. Institute of Content and Methods of Teaching, Russian Academy of Education, Moscow, Candidate of Pedagogical Sciences dissertation, in Russian.

Pan, Jennifer, Zijie Shao, and Yiqing Xu. 2022. "How Government-Controlled Media Shifts Policy Attitudes Through Framing." *Political Science Research and Methods*, 10(2): 317–332.

Portström, Fredrik. 2018. "parse_wiki_text: Parse wiki text from Mediawiki into a tree of elements." Rust crate, version 0.1.5.

Potapenko, Maxim, and Mateusz Kamionka. 2023. "Wiki-History of Crimea: Ukrainian and Russian Versions." In *Different Shades of the Past: History as an Instrument of Contemporary International Conflicts*. , ed. Mateusz Kamionka and Przemysław Łukasik, 91–104. Berlin and Boston:De Gruyter Oldenbourg.

Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda." *Journal of Economic Perspectives*, 31(1): 117–40.

Ratcliff, John W., and David Metzener. 1988. "Pattern Matching: The Gestalt Approach." *Dr. Dobb's Journal*, 46.

Reform.news. 2026. "Belarusian Wikipedia Warns Editors About Lack of Edit-Hiding Function." *Reform.news*, https://reform.news/en/belarusian-wikipedia-warns-editors-about-lack-of-edit-hiding-function, Accessed May 21, 2026.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall.* Princeton, NJ:Princeton University Press.

Rozhkov, Roman. 2025. "По всем статьям: ИИ поможет «Рувики» анализировать и верифицировать контент." *Forbes Russia*, Published September 18, 2025; accessed May 21, 2026.

Shadayev, Maksut. 2023. "Statement on Wikipedia blocking plans." Minister of Digital Development, Communications and Mass Media. As reported by *Interfax*, April 18, 2023. Available at: https://interfax.com/newsroom/top-stories/89721/.

Shadmehr, Mehdi, and Dan Bernhardt. 2015. "State Censorship." *American Economic Journal: Microeconomics*, 7(2): 280–307.

Steinsson, Sverrir. 2024. "Rule Ambiguity, Institutional Clashes, and Population Loss: How Wikipedia Became the Last Good Place on the Internet." *American Political Science Review*, 118(1): 235–251.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2022. "Why Botter: How Pro-Government Bots Fight Opposition in Russia." *American Political Science Review*, 116(3): 843–857.

Taddy, Matt. 2017. "One-Step Estimator Paths for Concave Regularization." *Journal of Computational and Graphical Statistics*, 26(3): 525–536.

The Signpost. 2024. "Ruwiki Is Run by Vladimir Medeyko, Wikipedia's Former Top Russian-Language Editor." *https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2024-04-25/In_the_media*, Accessed May 27, 2026.

Trokhymovych, Mykola, Oleksandr Kosovan, Nathan Forrester, Pablo Aragón, Diego Saez-Trumper, and Ricardo Baeza-Yates. 2025. "Characterizing Knowledge Manipulation in a Russian Wikipedia Fork." 1924–1936. AAAI Press.

Tsvetkova, Maria. 2025. "Avtomatizatsiya, II, proverka dannykh: kak izmenilas' "Ruviki" za dva goda." *Gazeta.Ru*, Accessed May 27, 2026.

Waight, Hannah, Eddie Yang, Yin Yuan, Solomon Messing, Margaret E. Roberts, Brandon M. Stewart, and Joshua A. Tucker. 2026. "State Media Control Influences Large Language Models." *Nature*.

Waight, Hannah, Solomon Messing, Anton Shirikov, Margaret E Roberts, Jonathan Nagler, Jason Greenfield, Megan A Brown, Kevin Aslett, and Joshua A Tucker. 2025. "Quantifying Narrative Similarity Across Languages." *Sociological Methods & Research*, 00491241251340080.

Wikimedia Foundation. 2022. "Wikimedia Foundation Appeals Russian Court Decision on Removal of Wikipedia Information Related to the Russian Invasion of Ukraine." *https://wikimediafoundation.org/news/2022/06/*

13/wikimedia-foundation-appeals-russian-court-decision-on-removal-of-wikipedia-information-related-to-the-russian-invasion-of-ukraine/.

Wikimedia Foundation Meta-Wiki. 2026. "Wikimedia Russia (Historical)." *https://meta.wikimedia.org/wiki/Wikimedia_Russia_(historical)*, Accessed May 27, 2026.

Wikipedia. 2024. "List of Controversial Issues." *https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues*, Accessed: 2025-01-09.

Wikipedia Signpost. 2025. "Two Years of Ruwiki: An Assessment." Available at: https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2025-06-24/Opinion.

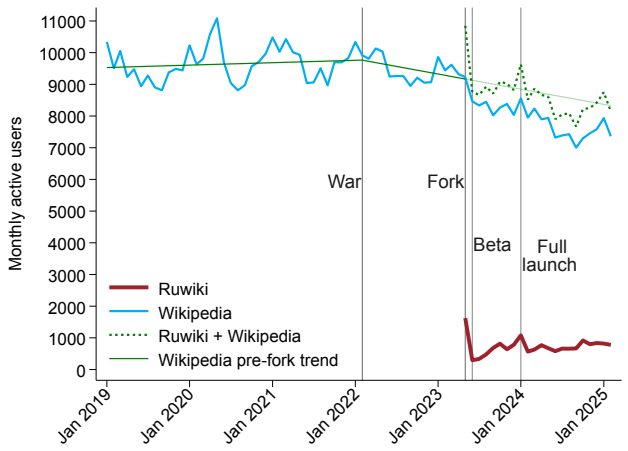
Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *The Quarterly Journal of Economics*, 129(4): 1947–1994.

Yurchak, Alexei. 2006. *Everything Was Forever, Until It Was No More: The Last Soviet Generation.* Princeton, NJ:Princeton University Press.

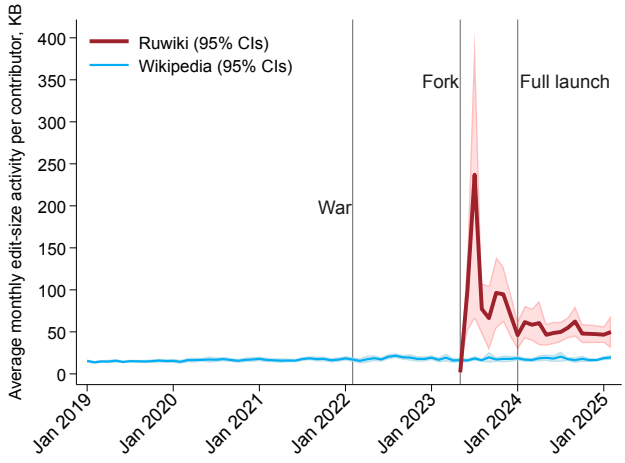
Радыё Свабода. 2025. "Вайна зь Вікіпэдыяй. Чаму ўлады Беларусі арыштоўваюць рэдактараў аднаго з самых папулярных у сьвеце сайтаў." *Радыё Свабода*, https://www.svaboda.org/a/33422252.html, Accessed May 21, 2026.

Figure 1: Contributors' activity

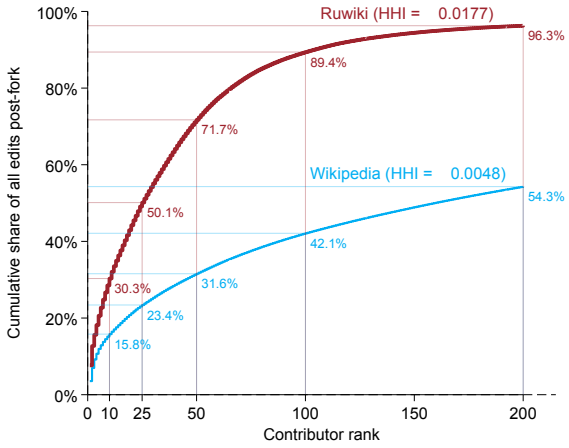
(a) Number of monthly active contributors



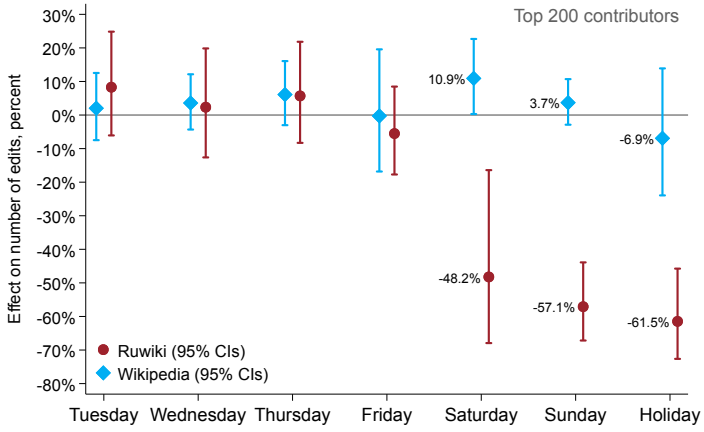
(b) Average monthly edit size per contributor, KB



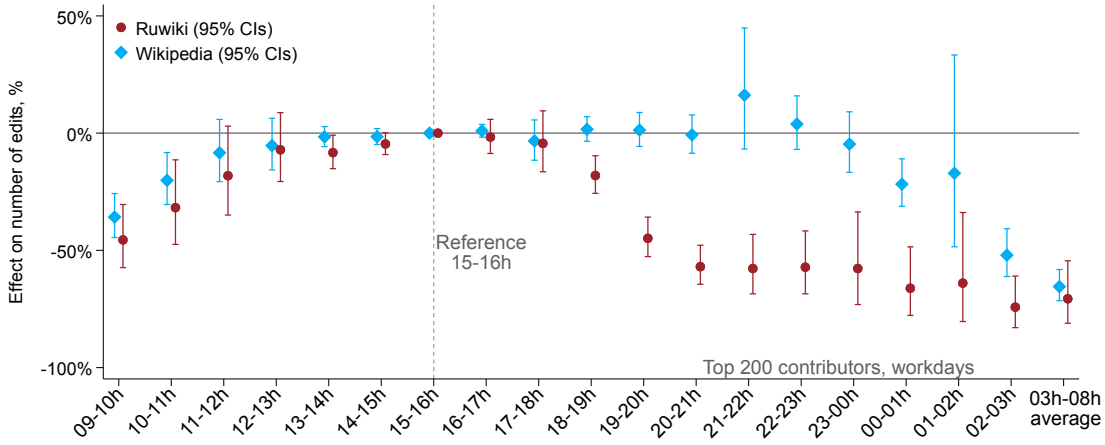
(c) Concentration in the number of edits, cumulative share for top 200 contributors



(d) Effect of type of day on number of edits, percentage change relative to ordinary Monday (PPML)



(e) Effect of hour of day on number of edits, percentage change relative to 15h-16h Moscow time (PPML)



Notes: The figure is based on analysis of 1,247,135 Ruwiki edits and 7,985,350 Wikipedia edits in the post-fork period, May 2023 to February 2025, as well as 19,292,169 Wikipedia edits pre-fork starting in Jan 2019. Panel A presents the number of contributors who made at least one edit by month. Panel B presents a proxy for the size of edits in KB per contributor per month, calculated as the sum of the absolute values of the changes to the length of the article as a result of the edit. Panel C presents the cumulative share of all post-fork edits made post fork by top contributors. Panels D and E report PPML estimates of how the number of edits made by the top 200 contributors varies across days of the week and holidays and across hours of the day. Regressions include contributor fixed effects. Coefficients are expressed as percentage changes.

Table 1: LLM Overall Narrative Assessment: Verbatim Model Output, Ruwiki vs. Wikipedia, Jan. 2025

All text below is reproduced verbatim from GPT-4.1 (temperature 0) and has not been edited by the authors.

Overall Assessment

The addition narratives reflect a concerted effort to legitimize Russian administrative, legal, and demographic control over disputed territories, reinforce anti-LGBTQ+ legislation, and frame military actions in official Russian terms. In contrast, deletions systematically remove language and data that support Ukrainian sovereignty, Western or opposition perspectives, LGBTQ+ content, and references to international sanctions. This pattern suggests a strong editorial push toward aligning Wikipedia content with official Russian narratives, minimizing dissent, and erasing Ukrainian and Western viewpoints from heavily edited articles.

Deletion Narratives

1. Removal of Invasion and Aggression Terminology: Deletions target language describing Russian actions as an 'invasion' or 'aggression,' as well as references to the war in Ukraine, shifting away from terminology that frames Russia as an aggressor.

2. Erasure of Ukrainian Sovereignty and Integrity References: Content referencing Ukraine's territorial integrity, sovereignty, and international support for Ukraine is removed, minimizing recognition of Ukrainian claims.

3. Suppression of Western and Opposition Sources: References to Western media, opposition figures, and critical sources are deleted, reducing visibility of dissenting or critical perspectives.

4. Censorship of LGBTQ+ and Sexual Content: Deletions remove references to LGBTQ+ identities, sexual orientation, and explicit content, aligning with Russian censorship policies and moral standards.

5. Omission of Sanctions and International Isolation: Mentions of sanctions, sanction lists, and international punitive measures against Russia or Russian individuals are systematically removed.

6. Removal of Ukrainian Administrative and Local Data: Deletions target Ukrainian administrative codes, local governance structures, and demographic data, erasing Ukrainian administrative presence from articles.

7. Erasure of Historical and Cultural Ukrainian References: Content related to Ukrainian history, culture, and notable figures is removed, diminishing the Ukrainian historical and cultural narrative.

Insertion Narratives

1. Integration of New Russian Federal Subjects: Additions emphasize the formal inclusion of new territories (e.g., Donetsk, Luhansk, Kherson, Zaporizhzhia) into the Russian Federation, referencing legal status, administrative changes, and official recognition.

2. Disputed Sovereignty and Ukrainian Contestation: Content is added to highlight ongoing disputes over territorial control, with references to Ukraine contesting Russian claims and continued assertions of Ukrainian sovereignty.

3. Legal and Administrative Reorganization: Narratives focus on the legal mechanisms and administrative restructuring of regions, including changes in laws, establishment or dissolution of settlements, and status updates.

4. Demographic and Census Data Updates: Additions include updated population statistics, gender breakdowns, and references to census data, reflecting efforts to document demographic changes in newly administered areas.

5. LGBTQ+ Legislation and Propaganda Restrictions: Content is added to reinforce Russian legal restrictions on LGBTQ+ rights, including bans on propaganda, same-sex marriage, and adoption, often contrasting with international norms.

6. Military Operations and War Framing: Additions frame Russian military actions as 'special military operations' and reference ongoing hostilities, often using official Russian terminology and avoiding terms like 'invasion.'

7. Administrative Geography and Local Governance: Narratives detail the reorganization of local governance, including the creation or merging of districts, settlements, and administrative units, often reflecting Russian administrative structures.

Notes: This table reproduces verbatim the output of the GPT-4.1 narrative-extraction pipeline with temperature set to 0. It retains the model's original style and formulations and should not be interpreted as author-written exposition. The authors have not edited the wording, narrative labels, or overall assessment. The cross-sectional analysis compares Wikipedia to Ruwiki, identifying systematic content differences. The analysis covers 56,835 deleted and heavily edited articles (54,049 heavily edited and 2,786 fully deleted, of which 2,783 yield text that enters the term analysis), comprising approximately 24.4 million deleted words (including the full text of the deleted articles, which are treated as deletions) and 11.4 million inserted words. The LLM identified seven deletion narratives and seven insertion narratives from the 500 terms with the highest χ^2 statistics in each direction, using bigrams and trigrams only.

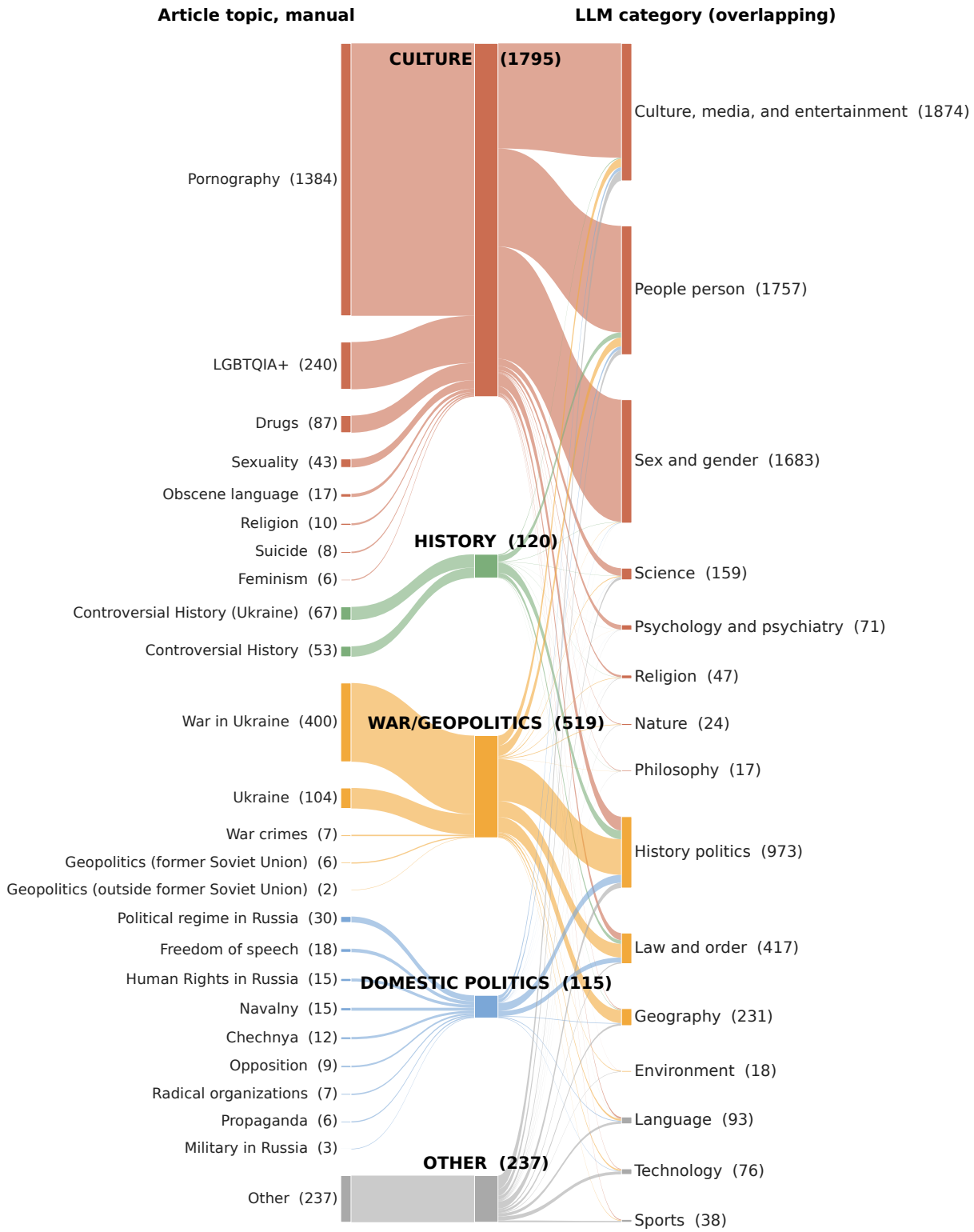
Table 2: LLM Narrative Assessments by Article Category: Verbatim Model Output, Ruwiki vs. Wikipedia, Jan. 2025

All text below is reproduced verbatim from GPT-4.1 (temperature 0) and has not been edited by the authors.

History & Politics (N = 28,107; 27,134 heavily edited, 973 deleted): The addition patterns reflect a strong push to legitimize Russian control over disputed territories, codify administrative changes, and assert Russian legal and cultural norms, especially in opposition to Ukrainian claims and Western values. Deletions systematically remove references to Russian aggression, Ukrainian sovereignty, international sanctions, and human rights issues, while also erasing Ukrainian administrative presence and critical media voices. This suggests a coordinated effort to reshape the historical and political narrative in favor of Russian state perspectives, while minimizing or erasing dissenting, Ukrainian, and international viewpoints.
Culture, Media & Entertainment (N = 21,842; 19,968 heavily edited, 1,874 deleted): The addition patterns reflect a strong push toward state-centric, nationalistic, and regulatory narratives, with increased focus on Russian legal frameworks, anti-LGBT and anti-Western stances, and the integration of war-related content. Simultaneously, deletions target Western adult industry references, explicit sexual content, critical reviews, and international or liberal perspectives, especially regarding Ukraine and LGBT rights. This suggests a concerted effort to align cultural and media content with official Russian narratives, emphasizing state control, traditional values, and the marginalization of dissenting or Western-influenced viewpoints.
Law & Order (N = 3,779; 3,362 heavily edited, 417 deleted): The addition patterns reflect a tightening of state-centric, conservative, and anti-LGBT narratives, emphasizing Russian legal sovereignty, traditional values, and the pathologization of gender and sexual diversity. Simultaneously, deletions systematically remove references to Russian aggression, human rights violations, LGBT rights, and independent or Western-aligned perspectives. This dual process constructs a narrative that legitimizes state policy, delegitimizes dissent and minority rights, and aligns legal and social discourse with official ideology, while actively erasing alternative or critical viewpoints from the record.
Sex & Gender (N = 3,720; 2,037 heavily edited, 1,683 deleted): The addition narratives reflect a strong trend toward legalistic, medicalized, and pathologizing framings of LGBT issues, with emphasis on restriction, criminalization, and the defense of traditional values. There is a clear effort to contrast Russian and non-Western stances with those of the West, often portraying LGBT rights as foreign or dangerous. In contrast, deletions systematically erase LGBT visibility, rights, and representation, especially in media, personal biographies, and legal contexts. This suggests a coordinated effort to both restrict and invisibilize LGBT topics, reinforcing conservative, exclusionary, and stigmatizing narratives.

Notes: This table reproduces verbatim the output of the GPT-4.1 narrative-extraction pipeline with temperature set to 0. It retains the model's original style and formulations and should not be interpreted as author-written exposition. The cross-sectional analysis compares Wikipedia to Ruwiki, identifying systematic content differences. The analysis covers 56,835 deleted and heavily edited articles (54,049 heavily edited and 2,786 fully deleted, of which 2,783 yield text that enters the term analysis), comprising approximately 24.4 million deleted words (including the full text of the deleted articles, which are treated as deletions) and 11.4 million inserted words. N indicates the number of articles in each category. The deletion side includes the full text of the fully deleted articles (treated as deletions), so N counts both groups: it is reported as the category total followed by its split into heavily edited and deleted articles. Assessments are based on the top 500 terms by χ^2 statistic in each direction (insertions and deletions). Only selected categories with corresponding visualizations in the main text are included.

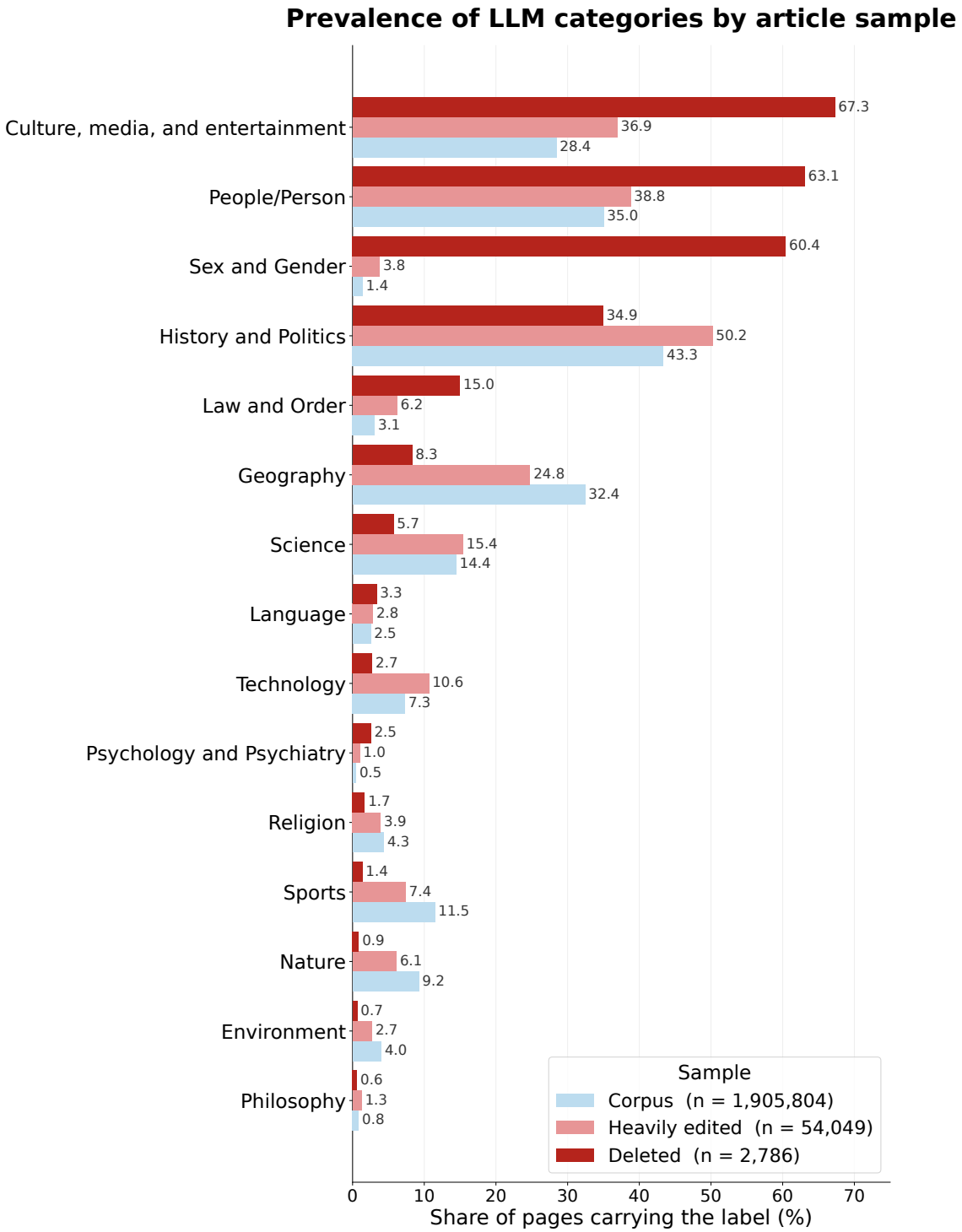
Figure 2: Topics of deleted articles: manual classification and LLM categories



Left/middle bar height = articles (mutually exclusive, one topic each); right bar height = LLM tag count (overlapping). Ribbons taper because LLM categories overlap -- an article may carry several tags. Nodes coloured by umbrella.

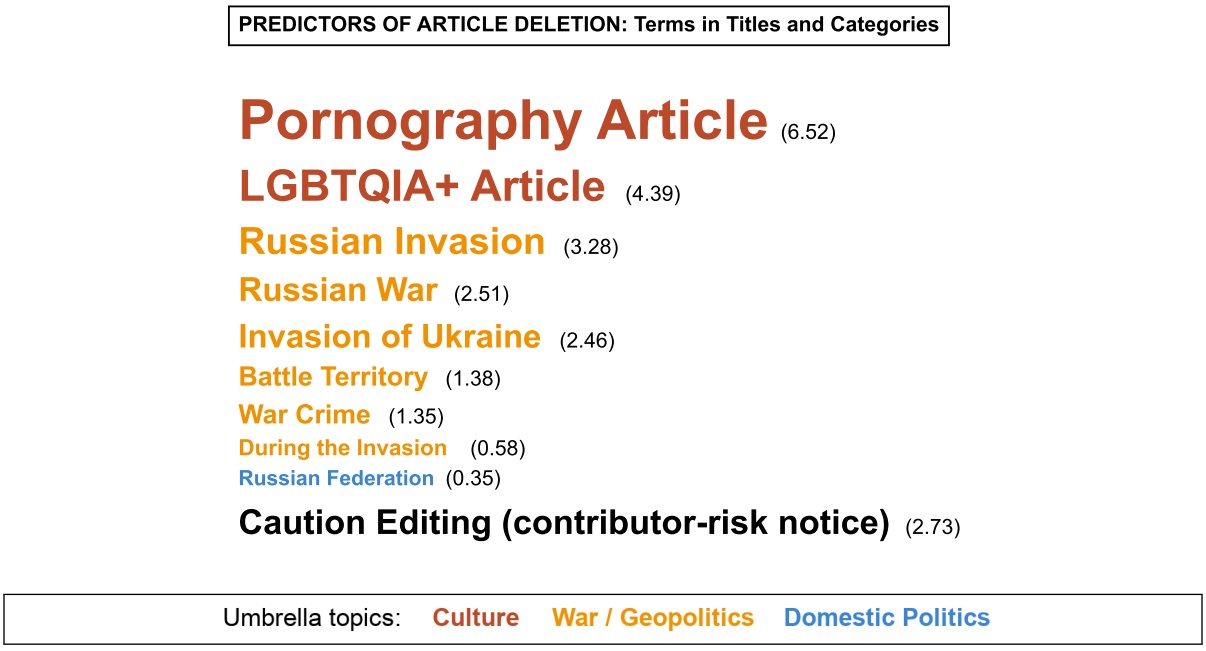
Note: The figure presents the topic categorization of all 2,786 deleted articles. It traces the flow of articles from mutually exclusive, fine-grained human-coded topics (left), through manually coded, mutually exclusive umbrella groups (center), to LLM-generated categories, which are not mutually exclusive (right). Ribbon widths on the left and in the center are proportional to the number of articles. The right-hand column is rescaled by a constant factor to match the total height of the left and center columns, while the labels report the actual number of articles in each LLM category.

Figure 3: LLM categories comparison: the entire corpus, deleted, and heavily edited articles



Note: The figure reports the share of articles in each sample assigned to each LLM category. Categories are not mutually exclusive: an article can be assigned to multiple categories, and therefore column percentages do not sum to 100. The full corpus consists of main-namespace Wikipedia articles used for the category analysis. Deleted articles are articles present in Wikipedia at the fork and on January 1, 2025 but absent from Ruwiki on January 1, 2025. Heavily edited articles are articles retained in Ruwiki but substantially modified relative to Wikipedia on January 1, 2025. For each category, we test whether category prevalence differs between samples using a series of category-specific logits. Tests are conducted separately by category. All differences for heavily edited articles are significant at $p < 0.001$. For deleted articles, all differences are significant at $p < 0.001$ except *language* ($p = 0.006$) and *philosophy* ($p = 0.173$).

Figure 4: LASSO predictors of deleted articles: Terms in Titles and Wikipedia Categories



Note: The figure presents results from logistic LASSO regression models predicting article deletion based on bigrams and trigrams extracted from article titles and Wikipedia category labels. The analysis is based on 100 bootstrap iterations, each comparing the 2,786 deleted articles to a fresh random sample of 5,000 non-deleted articles. All terms shown predict deletion ($\beta > 0$). Terms are ranked by β and set in a font size that increases with β; the size is indicative rather than proportional. For each term, the reported coefficient is the mean of its estimated LASSO coefficients across the bootstrap iterations in which it is selected, that is, iterations in which its coefficient is nonzero. Iterations in which the term is not selected are excluded from the average. The figure shows only terms with an estimated coefficient of at least 0.05 in at least 75% of the 100 iterations. Terms are translated from the original Russian into English.

Figure 5: LASSO predictions of deletions and insertions in heavily edited articles: Terms selected only in the baseline Ruwiki-Wikipedia comparison in January 2025

(a) All heavily edited
(49,926 articles, 838,219 segments)

DELETIONS: Ruwiki vs. Wikipedia (2025)

- Russian Invasion (-4.12)
- Invasion of Ukraine (-2.87)
- Russo-Ukrainian War (-2.26)
- Luhansk Oblast (occupied by Russia) (-2.22)
- Invasion by Russia (-2.11)
- Telephone Code (-1.96)
- Gay Lesbian (-1.76)
- Novaya Gazeta (opposition newspaper) (-1.65)
- Against Ukraine (-1.59)
- Alexei Navalny (opposition leader) (-1.57)

INSERTIONS: Ruwiki vs. Wikipedia (2025)

- Continue Consider Territory (+2.74)
- Region Territory of Ukraine (+2.43)
- Banned in Russian Federation (+2.25)
- Military Actions Ukraine (+2.10)
- Special Military Operation (+1.83)
- Kosovo Crisis (+1.80)
- Federation Defines (+1.71)
- New Subject (+1.64)
- Makeup of Ukraine (+1.60)
- LGBT Propaganda (+1.52)

(b) History and politics
(26,966 articles, 373,361 segments)

DELETIONS: History & Politics, Ruwiki vs. Wikipedia (2025)

- Russian Invasion (-3.94)
- Invasion of Ukraine (-2.99)
- Russo-Ukrainian War (-2.50)
- Invasion by Russia (-2.24)
- Luhansk Oblast (occupied by Russia) (-2.19)
- Gay Lesbian (-2.00)
- Novaya Gazeta (opposition newspaper) (-1.83)
- Against Ukraine (-1.71)
- Chuvashia, Russian (-1.65)
- Alexei Navalny (opposition leader) (-1.57)

INSERTIONS: History & Politics, Ruwiki vs. Wikipedia (2025)

- Continue Consider Territory (+3.27)
- Region Territory of Ukraine (+2.55)
- Banned in Russian Federation (+2.46)
- Military Actions Involving Ukraine (+2.41)
- Kosovo Crisis (+2.28)
- Special Military Operation (+2.03)
- Federation Defines (+1.84)
- Settlement Formation (+1.63)
- Makeup of Ukraine (+1.63)
- Duma Committee (+1.61)

(c) Law and order
(3,342 articles, 51,928 segments)

DELETIONS: Law & Order, Ruwiki vs. Wikipedia (2025)

- Russian Invasion (-4.16)
- Invasion of Ukraine (-2.89)
- Russo-Ukrainian War (-2.71)
- Novaya Gazeta (opposition newspaper) (-1.99)
- Intersex Rights (-1.97)
- Gay Lesbian (-1.89)
- Invasion by Russia (-1.85)
- Ministry of Justice of the Russian Federation (-1.81)
- Kosovo War (-1.69)
- Caution Editing (contributor-risk notice) (-1.65)

INSERTIONS: Law & Order, Ruwiki vs. Wikipedia (2025)

- Kosovo Crisis (+3.15)
- Banned in Russian Federation (+2.96)
- Many Countries World (+2.21)
- List of Designated Foreign Agents (+2.15)
- Intersex Registration (+2.06)
- Military Actions Involving Ukraine (+1.92)
- House of Soviets (+1.82)
- Person's Orientation (+1.77)
- Killing (+1.64)
- Same-sex Couples (+1.59)

(d) Sex and gender
(2,007 articles, 21,307 segments)

DELETIONS: Sex & Gender, Ruwiki vs. Wikipedia (2025)

- Russian Invasion (-2.92)
- Intersex Rights (-2.15)
- Gay Lesbian (-2.10)
- Ground of Hatred (-1.99)
- Legal Position (-1.88)
- Invasion of Ukraine (-1.88)
- Recognition of Same-sex Relationships (-1.80)
- Independent State (-1.78)
- Labia (-1.77)
- Sexual Intercourse (-1.71)

INSERTIONS: Sex & Gender, Ruwiki vs. Wikipedia (2025)

- Banned in Russian Federation (+3.94)
- Many Countries World (+2.71)
- Intersex Registration (+2.67)
- Transition Name (+2.44)
- Person's Orientation (+2.43)
- Course of Therapy (+2.40)
- Despite Protest (+2.10)
- LGBT Propaganda (+2.09)
- Same-sex Couples (+2.07)
- Famous Bearer (+2.04)

Umbrella topics: War / Geopolitics | Domestic Politics | Culture | History | Sport | Other

Notes: Terms selected by the LASSO in the January 2025 Ruwiki–Wikipedia comparison but not in the within–Wikipedia placebo. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. No terms shown are classified as HISTORY or SPORT; SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples: all heavily edited articles, 838,219 segments in 49,926 articles; history and politics, 373,361 in 26,966; law and order, 51,928 in 3,342; and sex and gender, 21,307 in 2,007.

Table 3: Text length and readability in heavily edited articles

	(1) *Length* Number of words	(2) *Sentence length* Words per sentence	(3) *Lexical complexity* Average word length	(4) *Readability* RU Flesch	(5) *Lexical diversity* MATTR-50 $\times 100$
Panel A: Specification without article fixed effects					
Ruwiki (vs Wikipedia)	-182.916***	-0.296***	0.039***	-0.447***	-0.100***
	(3.710)	(0.025)	(0.002)	(0.053)	(0.020)
Contributor-risk notice	36.195	-2.537***	0.174***	0.747	2.240***
	(44.073)	(0.160)	(0.011)	(0.475)	(0.172)
Ruwiki $\times$ Contributor-risk notice	-128.801***	2.313***	0.153***	-8.727***	-5.234***
	(26.973)	(0.120)	(0.010)	(0.420)	(0.209)
Panel B: Specification with article fixed effects					
Ruwiki (vs Wikipedia)	-182.916***	-0.296***	0.039***	-0.447***	-0.100***
	(3.710)	(0.025)	(0.002)	(0.053)	(0.020)
Ruwiki $\times$ Contributor-risk notice	-128.801***	2.313***	0.153***	-8.727***	-5.234***
	(26.973)	(0.120)	(0.010)	(0.420)	(0.209)
Panel C: Specification with article fixed effects and log number of words control					
Ruwiki (vs Wikipedia)		0.244***	0.001	-0.294***	-0.214***
		(0.030)	(0.002)	(0.061)	(0.023)
Ruwiki $\times$ Contributor-risk notice		1.982***	0.177***	-8.821***	-5.164***
		(0.107)	(0.010)	(0.423)	(0.206)
Wikipedia mean	782.767	13.409	6.326	30.652	84.816
Wikipedia SD	1352.905	6.855	0.521	17.285	5.736
Observations	108098	108098	108098	108098	108098
Articles	54049	54049	54049	54049	54049

Notes: The sample consists of heavily edited articles observed in both Wikipedia and Ruwiki. A contributor-risk notice serves as a proxy for politically sensitive content: it warns Wikipedia editors that contributing to the article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. Panels B and C absorb article fixed effects. Panel C additionally controls for the log number of words in the source-version text; its coefficient is not reported. Wikipedia mean and SD are calculated within the estimation sample for each outcome and are reported once at the bottom of the table. Standard errors are clustered by article and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Contradictions Between Official History Textbooks and Wikipedia vs. Ruwiki Articles

	(1)	(2)	(3)	(4)	(5)	(6)
	Any contradiction			Contradiction count		
Wikipedia (vs Ruwiki)	0.130***	0.128***	0.130***	1.139***	1.128***	1.151***
	(0.039)	(0.040)	(0.040)	(0.410)	(0.411)	(0.426)
Substantive claim counts		✓	✓		✓	✓
Article-pair FEs			✓			✓
Mean DV	0.125	0.125	0.125	0.165	0.165	0.786
Observations	200	200	200	200	200	42

Notes: Each observation is an encyclopedia article in either its Wikipedia or Ruwiki version matched to a textbook subchapter. As an outcome, we count direct contradictions between the official textbook subchapter and each encyclopedia article. The "Wikipedia (vs Ruwiki)" indicator equals one for the Wikipedia article and zero for the Ruwiki article. Columns 1–3 report OLS estimates with the binary outcome (Any contradiction); columns 4–6 report PPML estimates with the count outcome. Columns 2, 3, 5, and 6 control for total numbers of substantive claims both in the article and in matched textbook subchapter. Article-pair fixed effects are included in columns 3 and 6. Standard errors clustered by matched pair are reported in parentheses. PPML fixed-effect specifications omit matched pairs with zero total contradictions. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.10$.

Online Appendix

A Data: Article Sources and Data-Management Pipeline

This appendix documents our data-collection and management pipeline and describes the structure of the resulting datasets. By construction, the pipeline is identical for Wikipedia and Ruwiki at every stage for which the two platforms share the same infrastructure, enabling the matched cross-platform comparison that underlies our analysis.

A.1 Collection of Edit-History Data

Russian-language Wikipedia. We collect the complete edit history of the Russian-language Wikipedia from publicly available database dumps provided by the Wikimedia Foundation at https://dumps.wikimedia.org/. The dumps contain the revision history of pages in all namespaces and are distributed as large XML files. We process them using a two-stage architecture. First, we identify and catalog all dump files covering the period of interest. Second, we process these files in parallel on distributed cloud-computing infrastructure and convert them into columnar Parquet files optimized for analytical queries.[32] Within each page element, nested revision elements represent individual edits. Each revision contains the complete raw Wikitext of the page as it appeared immediately after that edit. We use containerized tasks to process dump files independently and in parallel, reducing the processing time for the full Russian-language Wikipedia from days to hours. The resulting data are stored in cloud object storage (Amazon S3) and made queryable through partitioned tables in Amazon Athena.

The resulting edit-history dataset for the Russian-language Wikipedia contains one record per revision and twenty-one fields describing each edit. Table A1 lists and defines these fields.

Ruwiki. Ruwiki does not publicly release database dumps. We therefore collect its data through an application programming interface (API) based on the same MediaWiki specification used by Wikipedia. Because Ruwiki retained Wikipedia's MediaWiki-based infrastructure when it forked the encyclopedia, its API provides broadly comparable access to article content and revision histories, allowing us to construct parallel datasets for the two platforms. Data collection proceeds in two stages. First, we query the API iteratively to retrieve the complete list of Ruwiki articles, including their identifiers and titles. Second, for each article, we retrieve its current content and complete revision history. We use pagination to recover the full histories of articles with large numbers of revisions and automated retries to address transient network errors.

Data quality assurance. We use several validation procedures to assess the completeness and consistency of the collected data. For Wikipedia, we compare the numbers of pages and revisions in the processed data with the corresponding counts reported in Wikimedia dump metadata. For Ruwiki, we maintain processing metadata recording which articles have been collected and when each record was last updated. These metadata allow us to identify gaps in coverage and perform incremental updates. The collection pipeline also handles platform-specific edge cases, including redirects, restricted or unavailable content, and, for Ruwiki, articles with revision histories that require multiple paginated API requests.

[32] We parse the XML files using a streaming parser that processes elements sequentially rather than loading entire files into memory, allowing us to handle multi-gigabyte files efficiently.

Table A1: Wikipedia Dump File Variables: Complete Field Specification

Field	Description	Category
Database & File Tracking		
wiki_db	Database name (e.g., "ruwiki")	Tracking
source_file	Original dump filename	Tracking
file_identifier	File partition identifier (e.g., "p1", "stub1")	Tracking
Page Metadata		
page_id	Unique article identifier	Page
page_title	Page title	Page
page_ns	Namespace (0 = main article, 1 = talk, etc.)	Page
page_redirect	Redirect target (if article is a redirect)	Page
Revision Metadata		
id	Unique revision identifier	Revision
parentid	Parent revision ID (previous version)	Revision
timestamp	Edit timestamp (ISO 8601 format)	Revision
minor	Minor edit flag ("true" if minor, empty otherwise)	Revision
comment	Edit summary/comment	Revision
Content Metadata		
model	Content model (typically "wikitext")	Content
format	Content format (typically "text/x-wiki")	Content
sha1	SHA-1 hash of content	Content
text_bytes	Size of complete article in bytes[a]	Content
text_id	Text identifier reference	Content
Full Content (varies by dump type)		
text	Full Wikitext content of article[b]	Content
Contributor Information		
contributor_username	Editor username (if registered user)	Contributor
contributor_id	Editor user ID (if registered user)	Contributor
contributor_ip	Editor IP address (if anonymous edit)	Contributor

[a] The total size of the complete article content after this revision in bytes – not the size of the change/delta.

[b] Contains the complete Wikitext snapshot of the article as it appeared after this revision. Empty in stub-meta-history files; populated only in complete history dump files.

Temporal coverage. The Russian-language Wikipedia dataset contains the complete revision history from the encyclopedia's inception through February 2025. For articles present on Ruwiki at the fork, the Ruwiki dataset contains all post-fork revisions through February 2025; for articles created on Ruwiki after the fork, it contains all revisions from their creation through February 2025. These data allow us to study editing activity on Wikipedia before the fork, the state of both platforms at the fork in May 2023, and their subsequent evolution. Our analysis of editing activity covers January 2019 through February 2025, while our analysis of content differences compares the two platforms in January 2025, twenty months after the fork.

A.2 Wikitext Parsing and Identification of Textual Changes

Cross-sectional pairing. We begin with the set of articles present in the Russian-language Wikipedia at the fork on May 1, 2023. For each article, we identify the most recent revision available on each platform as of January 1, 2025. The resulting paired dataset contains the article identifier, title, revision timestamp, and full Wikitext of the Wikipedia and Ruwiki versions. Articles that remain present on Wikipedia but are absent from Ruwiki as of January 1, 2025 are classified as deletions; articles present on both platforms constitute the sample used for content comparisons. For the within-Wikipedia placebo analyses, we construct an analogous pair of snapshots for the same article universe, using the most recent Wikipedia revision as of May 1, 2023 and January 1, 2025.

Wikitext parsing. Before comparing article content, we convert raw Wikitext into plain text using a custom Rust-based parser built on the `parse_wiki_text` library (Portström, 2018). The parser removes MediaWiki markup and non-prose elements, including template invocations, infoboxes, tables, reference tags, and link syntax, while retaining the main encyclopedic text. Recursive processing allows it to handle nested templates and other complex markup structures. Parallelization across multiple CPU cores yields a processing rate of approximately one thousand articles per minute. To prevent unusually complex articles from stalling the pipeline, we impose a thirty-second processing limit per article. Articles that exceed this limit are skipped automatically and logged for review. For each matched article, the parser produces two plain-text files, one for the Wikipedia version and one for the Ruwiki version. These files are used for automated detection of textual differences and for manual inspection.

Identification and visualization of textual differences. Using the cleaned text for the two versions of each article, we identify insertions and deletions with Python's `difflib.SequenceMatcher`. The algorithm implements a variant of the Ratcliff–Obershelp pattern-matching approach (Ratcliff and Metzener, 1988), recursively identifying the longest contiguous matching blocks in the two texts. We classify matching blocks as unchanged text, unmatched text appearing only in Wikipedia as deletions from Ruwiki, and unmatched text appearing only in Ruwiki as additions to Ruwiki. We then generate HTML files that display these differences using color-coded highlighting. These files facilitate manual validation of the automated comparisons and provide the visual evidence presented in the case studies in Online Appendix J.

Heavy-edit threshold. For each article, we calculate the numbers of characters deleted from and added to the original Wikipedia version, as well as their sum. We normalize each measure by the character count of the original Wikipedia version. Among articles containing at least 100 words, we classify an article as *heavily edited* if the deleted text accounts for more than 10 percent of the original article length, the added text accounts for more than 10 percent, or the combined volume of added and deleted text exceeds 15 percent.

A.3 Cleaning of Textual Differences and N-gram Extraction

The structured panel of text added to and deleted from each article serves as the input to all term-level analyses in Sections V and VI.B. Russian morphology poses particular challenges because nouns are extensively declined and verbs are inflected for tense, aspect, and mood. We process the text using spaCy (Honnibal et al., 2020) with a Russian-language model and use pymorphy2 (Korobov, 2015) as a fallback when spaCy does not provide a satisfactory morphological analysis.

The cleaning pipeline removes residual Wikipedia-specific markup, HTML artifacts, reference markers, and terms associated with Wikipedia's technical infrastructure. It then normalizes whitespace and removes single-character tokens, numeric strings, and Russian- and English-language stopwords. Based on manual inspection of the resulting text, we also exclude a fixed list of 270 unique terms consisting almost entirely of residual MediaWiki markup. These include template invocations, among them 96 terms containing the Russian token for "Template:"; citation and HTML fragments; maintenance-tag phrases; and high-frequency layout boilerplate from sports tables, artwork infoboxes, and population figures. Much of this boilerplate reflects Ruwiki's redesign of the corresponding infoboxes. No political or narrative vocabulary is excluded. A small number of excluded template names refer to politically salient templates, such as the "foreign agent" designation template. However, these are template invocations rather than article prose, and excluding them does not alter the article text. The complete annotated list is available in the replication package. We use the resulting cleaned and lemmatized text to construct the n-grams employed in the subsequent statistical analyses.

We construct binary document–term matrices recording the presence of bigrams and trigrams in each text segment. To focus on content-bearing expressions, we retain only n-grams that contain at least one noun. We further require each n-gram to appear in at least 5 but no more than 5,000 distinct text segments, out of 896,973 pooled insertion and deletion segments, and in at least 4 distinct articles. The upper threshold corresponds to approximately 0.6 percent of all segments. All frequency filters and association statistics are based on segment-level presence indicators rather than raw occurrence counts. Only two n-grams exceed the upper threshold: "российский федерация," the lemmatized form of "Russian Federation," which appears in 9,449 segments, and "настоящий время," the lemmatized form of "at present," which appears in 5,475 segments. Both are near-ubiquitous boilerplate expressions. These restrictions exclude rare, potentially noisy expressions and highly frequent boilerplate while preserving terms with sufficient variation for statistical analysis. We construct separate matrices for insertions and deletions, both for the full sample and within each topic category, enabling the category-specific analyses reported in Section V.

B Article Classification with DeepSeek-V3

To characterize the thematic distribution of editorial intervention, we use the DeepSeek-V3 large language model (DeepSeek-AI et al., 2024) to classify 1.9 million main-namespace articles in the Russian-language Wikipedia into seventeen substantive topic categories.[33] For each article, the model receives the title and associated Wikipedia category labels and returns seventeen binary indicators, one for each topic category. Our classification scheme is adapted from Wikipedia (2024), which identifies broad topic areas commonly associated with editorial disputes on Wikipedia. We modify the original scheme by separating several related topics—including Culture and Entertainment, Media, and Sports—and by distinguishing Nature and Geography from Environment and Science, thereby providing more comprehensive coverage of the corpus.

The resulting categories are History, Politics, Religion, Science, Culture and Entertainment, Sex and Gender, Environment, Law and Order, Language, Philosophy, Psychology and Psychiatry, Technology, Media, Sports, People/Person, Nature, and Geography. The classification is non-exclusive, so an article may be assigned to multiple categories. The full classification prompt is reproduced below.

LLM Prompt for Article Categorization

```
You are an expert in media studies and your job is to read the title and categories
of russian wikipedia articles and assign them to the following classes:
'History, Politics, Religion, Science, Culture and Entertainment, Sex and Gender,
Environment, Law and Order, Language, Philosophy, Psychology and Psychiatry,
Technology, Media, Sports, People/person, Nature, Geography'

For each of them individually, decide whether the article belongs to the class or
not by returning a 1 for 'yes' and a 0 for 'no'. In other words, you can select
multiple classes for each article.
Return your answer in the following JSON format, follow it exactly:
{
  "history": int,
  "politics": int,
  "religion": int,
  "science": int,
  "culture_and_entertainment": int,
  "sex_and_gender": int,
  "environment": int,
  "law_and_order": int,
  "language": int,
  "philosophy": int,
  "psychology_and_psychiatry": int,
  "technology": int,
  "media": int,
  "sports": int,
  "people_person": int,
  "nature": int,
  "geography": int,
```

[33] Recent evidence shows that GPT-based classifications perform well against human labels and are robust to semantically equivalent prompt formulations (Asirvatham, Mokski and Shleifer, 2026).

```
}

Here is the title and categories of the article:

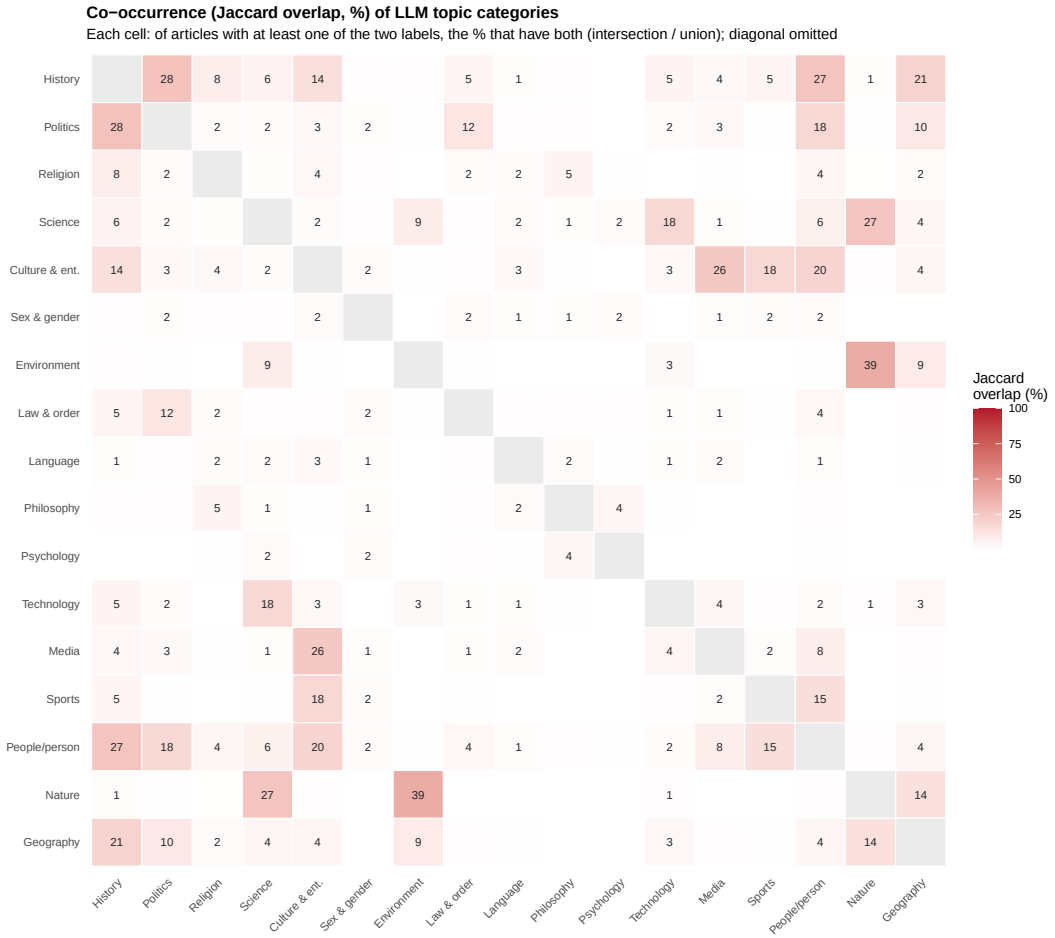
"[ARTICLE_TITLE -- CATEGORY_LABELS]"

Your answer:
```

We run the classification pipeline in parallel using 100 concurrent API workers and periodically save intermediate results. The pipeline is resumable: upon restarting, it identifies and skips articles that have already been classified successfully. We require model responses to conform to the specified structured-output format. Responses that are invalid or cannot be parsed are flagged and excluded from the final classification dataset. Valid classification labels are assigned to 1,905,804 articles.

The model was asked to assign each article to seventeen binary categories, with history coded separately from politics and culture & entertainment coded separately from media. After completing the LLM classification, we combine history and politics into a single category and culture & entertainment and media into another single category. The subsequent analysis therefore uses fifteen categories. Appendix Figure A1 presents the category co-occurrence among heavily edited articles.

Figure A1: LLM category co-occurrence among heavily edited articles

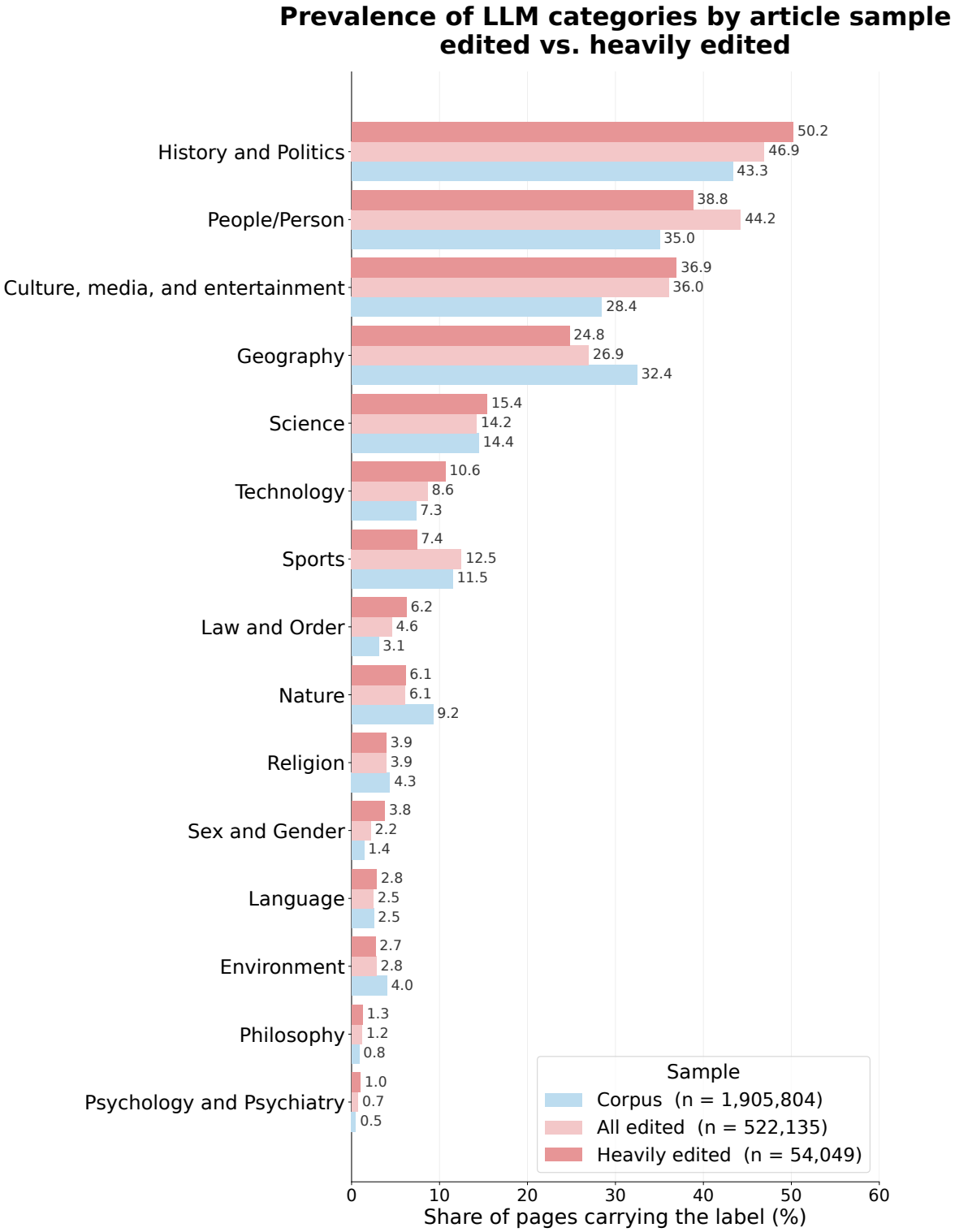


Note: We report the Jaccard overlap, in percent, for each cell, computed over the 17 binary LLM category labels; the diagonal, which equals 100% by construction, is omitted. To read a cell, consider all articles that have at least one of the two category labels. The reported percentage is the share of those articles that have both labels:

$$J_{ij} = 100 \times \frac{|i \cap j|}{|i \cup j|}.$$

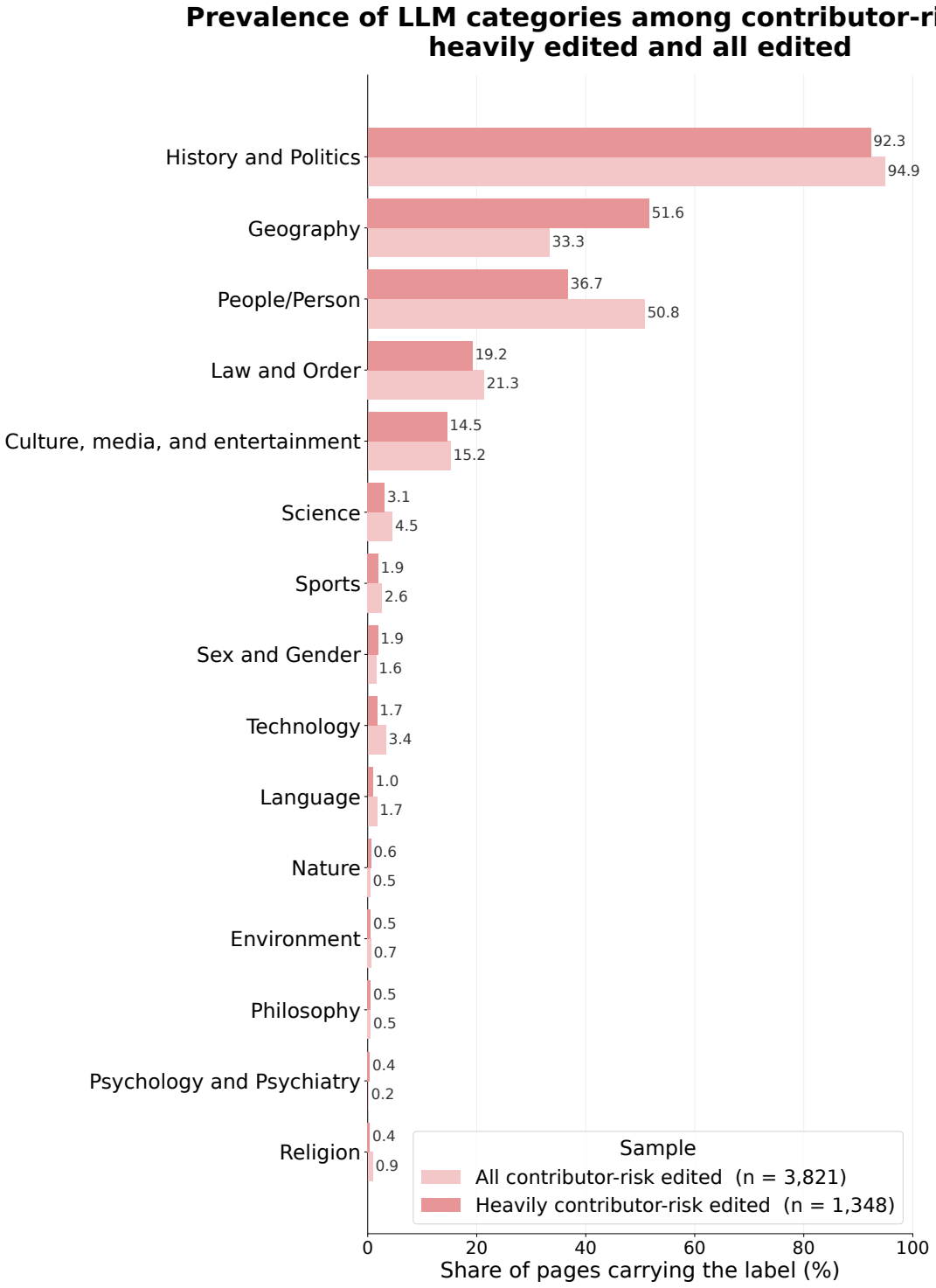
The measure ranges from 0%, when no article has both labels, to 100%, when the two labels always occur on exactly the same articles. Higher values indicate greater overlap between the two categories. Because labels are not mutually exclusive, this is an overlap measure, not a conditional probability or a correlation.

Figure A2: LLM categories comparison: the entire corpus, edited, and heavily edited articles



Note: The figure reports the share of articles in each sample assigned to each LLM category. Categories are not mutually exclusive: an article can be assigned to multiple categories, and therefore column percentages do not sum to 100. The full corpus consists of main-namespace Wikipedia articles used for the category analysis. Heavily edited articles are articles retained in Ruwiki but substantially modified relative to Wikipedia on January 1, 2025. Edited articles are all articles retained in Ruwiki and modified relative to Wikipedia on January 1, 2025.

Figure A3: LLM categories: articles that carry contributor-risk notice among heavily edited and all edited



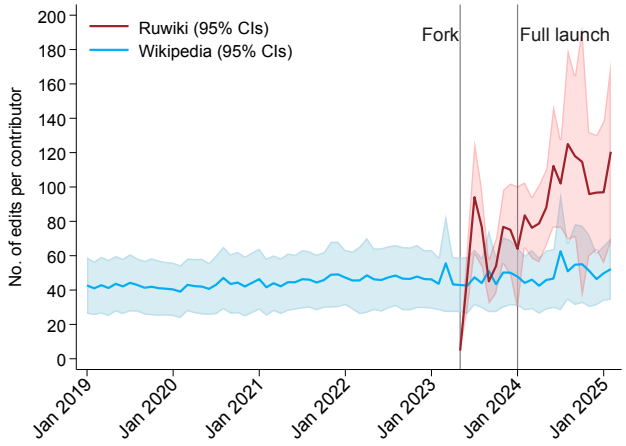
Note: The figure reports the share of articles in each sample assigned to each LLM category. Categories are not mutually exclusive: an article can be assigned to multiple categories, and therefore column percentages do not sum to 100. The samples are heavily edited articles that carry contributor-risk notice and all edited articles that carry contributor-risk notice.

C Editing Activity: Additional Figures

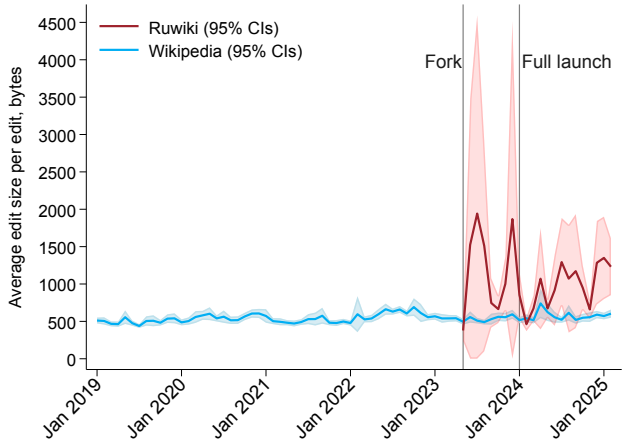
This appendix section provides supplementary evidence on the editing-activity patterns documented in Section IV.

Figure A4: Averages of the number of edits, size per edit, and size of monthly additions and deletions

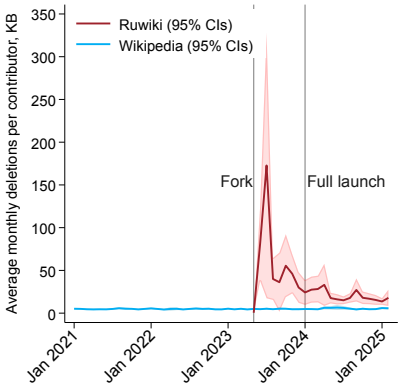
(a) Number of monthly edits per contributor



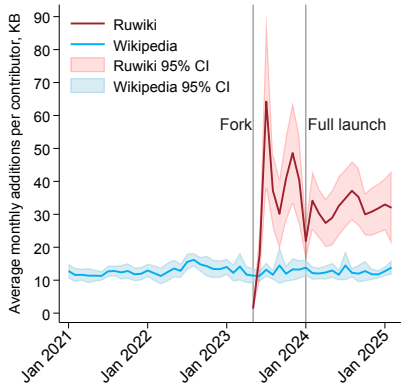
(b) Average edit size per edit and per contributor, bytes



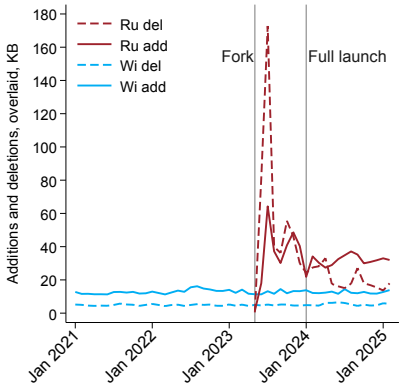
(c) Size of net deletions



(d) Size of net additions



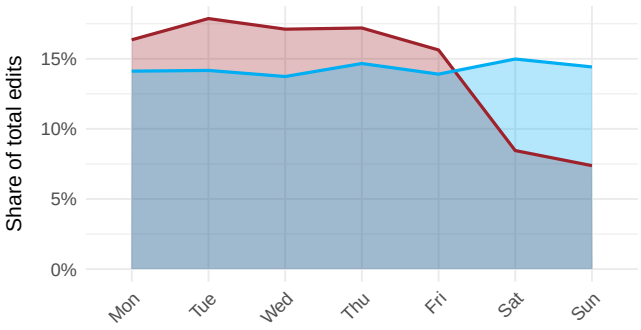
(e) Additions and deletions overlaid



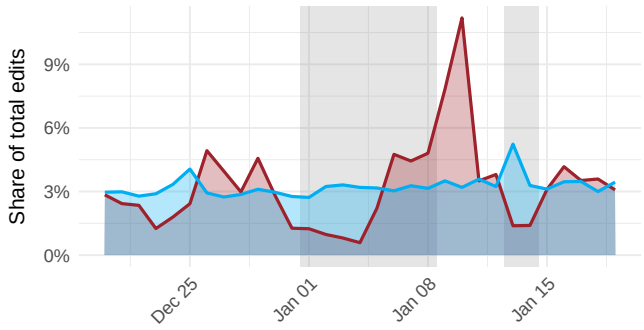
Notes: The figure is based on analysis of 1,247,135 Ruwiki edits and 7,985,350 Wikipedia edits in the post-fork period, May 2023 to February 2025, as well as 19,292,169 Wikipedia edits pre-fork starting in Jan 2019. Panel A presents the number of monthly edits per contributor. Panel B presents a proxy for average edit size, calculated within each contributor-month as the mean absolute change in article length per edit. Panels C–E present the sizes of net deletions and net additions.

Figure A5: Raw-data: Daily and hourly distribution of contributors' activity

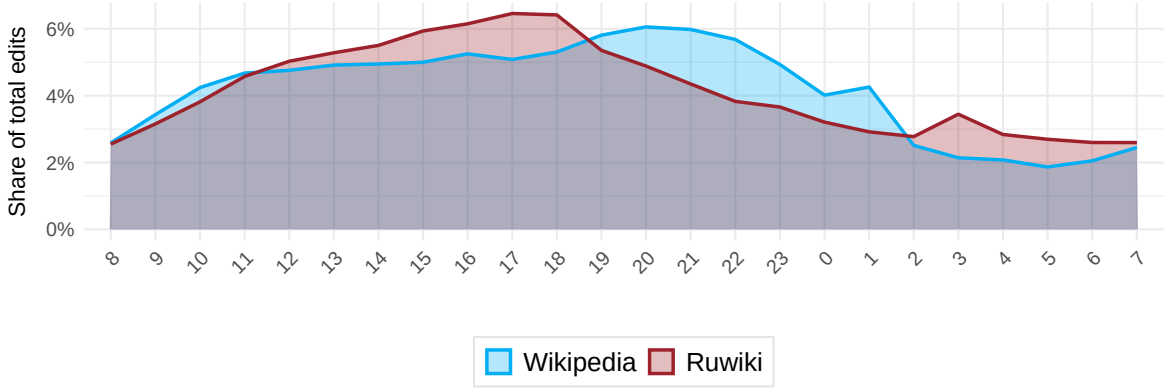
(a) Activity by day of week and holidays



(b) Activity during the New Year holidays



(c) Activity by hour of day (Moscow time)



Notes: This figure compares the editing patterns (excluding bots) on Ruwiki and Wikipedia. It shows the distribution of edits by day of week (Panel A), around New Year holidays (Panel B), and by hour of day in Moscow time (Panel C) for top the 200 contributors (excluding bots). The values are expressed as shares normalized by total edits to highlight relative activity patterns. An active contributor is a distinct user with at least one edit in a given period.

D Narrative-Identification Prompts and Robustness

This appendix documents the LLM prompts used for narrative identification and validation in Section V.

D.1 Narrative-Identification Prompt

The first prompt identifies distinct narratives from the top 500 terms most strongly associated with additions and deletions. This prompt is used in the three-step methodology described in Section V, Step 2. The model (GPT-4.1 with temperature 0) receives lists of Russian terms and synthesizes them into coherent narratives.

LLM Prompt for Narrative Identification

```
Developer Message:

You are an expert in content analysis and narrative identification. You will
analyze terms that are statistically associated with additions and deletions in
Wikipedia articles within specific categories.

Your task is to identify potential narratives reflected in these patterns of
additions and deletions.

For each category, you will receive:
1. Terms strongly associated with ADDITIONS (content being added to articles)
2. Terms strongly associated with DELETIONS (content being removed from articles)

Please provide your analysis in a structured JSON format with the following
structure:
{
  "category": "[category_name]",
  "addition_narratives": [
    {
      "narrative": "[NARRATIVE_TITLE_1]",
      "explanation": "[CONCISE_EXPLANATION_1_MAX_100_WORDS]",
      "key_terms": ["term1", "term2", "term3"]
    },
    {
      "narrative": "[NARRATIVE_TITLE_2]",
      "explanation": "[CONCISE_EXPLANATION_2_MAX_100_WORDS]",
      "key_terms": ["term4", "term5", "term6"]
    }
    // Include exactly {narrative_count} addition narratives
  ],
  "deletion_narratives": [
    {
      "narrative": "[NARRATIVE_TITLE_1]",
      "explanation": "[CONCISE_EXPLANATION_1_MAX_100_WORDS]",
      "key_terms": ["term1", "term2", "term3"]
    },
    {
```

```
      "narrative": "[NARRATIVE_TITLE_2]",
      "explanation": "[CONCISE_EXPLANATION_2_MAX_100_WORDS]",
      "key_terms": ["term4", "term5", "term6"]
    }
    // Include exactly {narrative_count} deletion narratives
  ],
  "overall_assessment": "[CONCISE_OVERALL_ASSESSMENT_MAX_200_WORDS]"
}

Identify {narrative_count} distinct narratives for both additions and deletions.
Be specific and base your analysis directly on the terms provided.

IMPORTANT: Keep all explanations concise -- maximum 100 words per narrative
explanation and maximum 200 words for the overall assessment.

User Message:

Category: {category}

TERMS ASSOCIATED WITH ADDITIONS:
{additions_text}

TERMS ASSOCIATED WITH DELETIONS:
{deletions_text}
```

D.2 Narrative-Relationship Comparison Prompt

The second prompt validates the hierarchical structure of identified narratives by comparing how narratives at different granularity levels relate to each other (Section V, Step 3). When we request different numbers of narratives (1 through 8), this prompt determines whether coarser narratives split into finer-grained ones or whether entirely different narratives emerge.

LLM Prompt for Narrative Relationship Analysis

```
Developer Message:

You are an expert in content analysis and narrative comparison. You will analyze
how narratives extracted at different levels relate to each other.

I will provide you with two sets of narratives in JSON format:
1. Lower level: Narratives extracted when instructed to identify fewer narratives
2. Higher level: Narratives extracted when instructed to identify more narratives

Your task is to analyze the relationship between these narratives and determine
how the narratives at the higher level relate to those at the lower level. For each
pair of narratives, choose the SINGLE MOST APPROPRIATE relationship type from the
following:

- CONTAINED: A higher-level narrative is fully contained within a lower-level
```

```
narrative
- SPLIT: A lower-level narrative has been split into multiple higher-level
narratives
- NEW: A higher-level narrative introduces content not present in any lower-level
narrative

Please provide your analysis in a simplified graph representation format as follows:

{
  "edges": [
    {
      "source": "L1_N1", // Unique ID of the lower-level narrative (format:
L[level]_N[narrative_id])
      "target": "L2_N1", // Unique ID of the higher-level narrative (format:
L[level]_N[narrative_id])
      "type": "RELATIONSHIP_TYPE", // ONE relationship type from the list above
      "weight": 0.8 // A value between 0 and 1 indicating the strength of the
relationship
    },
    // Additional edges as needed
  ]
}

IMPORTANT: Each target (higher-level narrative) should appear only ONCE for each
source (lower-level narrative). Choose the single most appropriate relationship
type for each source-target pair.

Please include narrative titles for both source and target in your output JSON using
keys 'source_title' and 'target_title'. Focus only on the relationship structure
between narrative IDs. Use the unique_id values provided in the input data for
source and target.

User Message:

{input_json} // Contains structured JSON with category, direction, and narratives
at both levels
```

D.3 Technical Specification for Assessing Hierarchical Stability

To assess stability across levels of granularity, we compare the narratives identified at each level with those identified at the immediately preceding, coarser level. Algorithm 1 formalizes this procedure. In a single query, GPT-4.1, with temperature set to 0, receives the complete sets of narratives from both levels and classifies the cross-level relationship of each finer-level narrative as *contained*, *split*, or *new*. A narrative is classified as *contained* if it falls within a broader narrative identified at the coarser level, as *split* if it is one of several finer narratives into which a coarser narrative divides, and as *new* if it has no substantive counterpart at the coarser level. Each proposed link receives exactly one of these three labels. For the visualization, we additionally derive a residual *continuation* status. When the strongest link of a finer-level narrative is labeled *split*, but its coarser-level parent has no other finer-level children, we treat the narrative as a continuation rather than a split. Although a *new* narrative has no substantive

counterpart at the coarser level, the edge list still assigns it a nominal source narrative, corresponding to the model's closest coarser-level anchor. This link determines only where the new branch is placed in Online Appendix Figure A6; it does not affect the narrative's classification.

The model also assigns each proposed cross-level link a confidence weight $w \in [0, 1]$. Links with $w > 0$ are retained as edges in the resulting directed graph. In the visualization, we label both *contained* and *continuation* relationships as *still present*, since both indicate continuity with a narrative identified at the coarser level. We display *split* relationships separately and label narratives with no coarser-level counterpart as *new*. The resulting sunburst visualizations are presented in Online Appendix Figure A6.

Algorithm 1: Compare narratives between two levels

Input: Lower level $\mathcal{L}$, Higher level $\mathcal{H}$
Output: Edge list $\mathcal{E}$
$\mathcal{E}' \leftarrow$ GPT-4.1$(\mathcal{L}, \mathcal{H})$; `// single call; returns typed, weighted edge list`
$\mathcal{E} \leftarrow \{\langle n_l, n_h, type, w\rangle \in \mathcal{E}' : w > 0\}$, deduplicated on (n_l, n_h) keeping the first occurrence;
return $\mathcal{E}$

E Conceptual Framework for the Baseline Ruwiki–Wikipedia and Placebo Within-Wikipedia Comparisons

In this appendix section, we develop a simple conceptual framework that guides our comparisons of Ruwiki and Wikipedia content.

Setup. Let F denote the fork date and let T denote the baseline comparison date, January 1, 2025. Let W denote the Wikipedia corpus and R the Ruwiki corpus. We consider four snapshots: Wikipedia at the fork (W_F), Wikipedia at the baseline date (W_T), Ruwiki at the fork (R_F), and Ruwiki at the baseline date (R_T). Because Ruwiki was created as a copy of Wikipedia, the two corpora coincide at the fork: $R_F = W_F$ by construction.

Between the fork and January 2025, the content of the two online encyclopedias could evolve through three conceptually distinct processes.

Three processes.

P1: **Change in the world between F and T.** Events occur, generating new facts, new content, and new vocabulary for an encyclopedia to cover.

P2: **Encyclopedic updating versus staleness.** Editors revise encyclopedic content to reflect changes in the world, as well as routine improvements in presentation, wording, sourcing, and structure. We assume that Wikipedia updates continuously to reflect current encyclopedic knowledge. Ruwiki may incorporate Wikipedia's updates fully, partially, or not at all. We refer to the failure to incorporate relevant updates as staleness or censorship by omission.

P3: **Active Ruwiki editing.** Ruwiki's editors may deliberately alter inherited content beyond ordinary updating by removing material—censorship—or inserting material—propaganda. This is the object of our interest.

The frozen-clone benchmark. It is useful to consider an extreme benchmark in which Ruwiki's editors make no edits after the fork. In this frozen-clone benchmark, Ruwiki remains equal to the fork-date Wikipedia snapshot, so that $R_T = R_F = W_F$. Any difference between R_T and W_T would then be mechanically identical to the difference between W_F and W_T, but with the opposite sign. The terms added to Wikipedia between F and T will appear missing from R_T as compared to W_T. Thus, under pure staleness, the cross-platform comparison (R_T vs. W_T) and the within-Wikipedia comparison (W_T vs. W_F) would reveal the same changes with the opposite sign.

The two comparisons: baseline R_T vs. W_T and placebo W_T vs. W_F. Throughout the content analysis, we present two comparisons. The baseline comparison is the cross-platform comparison between R_T and W_T. The placebo comparison is the within-Wikipedia comparison between W_T and W_F.

These two comparisons isolate different combinations of the processes above. The within-Wikipedia comparison captures how Wikipedia changed between the fork and January 2025. It therefore reflects changes in the world as incorporated by Wikipedia's updating, together with routine editorial drift. By construction, it contains no trace of Ruwiki's behavior.

The cross-platform comparison, in contrast, compares the two corpora at the same date. It therefore holds the state of the world fixed and captures the divergence between Ruwiki and Wikipedia by January 2025. This divergence can arise from two sources: Ruwiki's staleness or censorship by omission, meaning the portion of Wikipedia's updating that Ruwiki did not incorporate, and active Ruwiki editing, meaning censorship and propaganda. A lack of updating in Ruwiki may itself reflect censorship if editors selectively leave damaging information outdated while incorporating neutral or favorable developments.

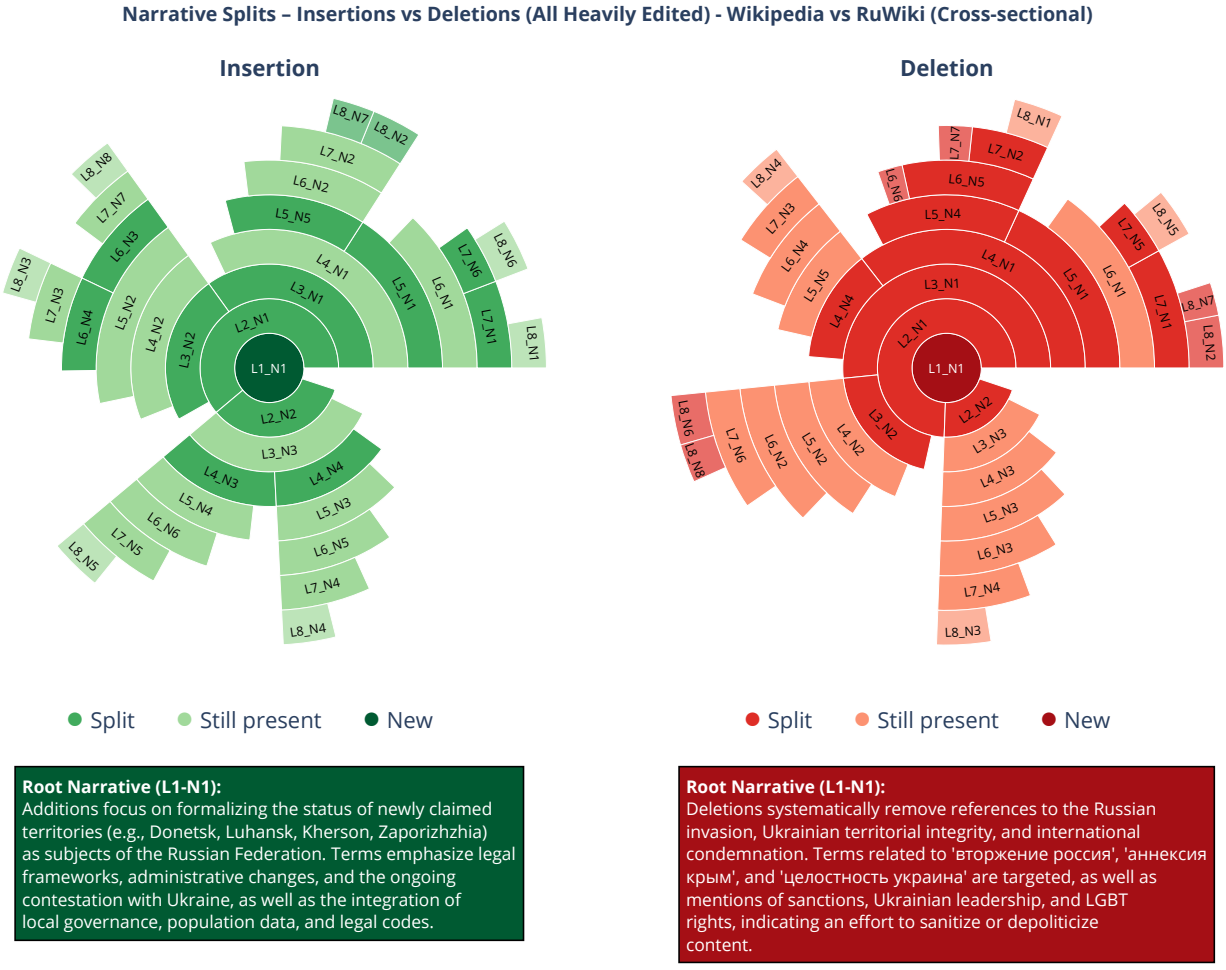
Appendix Section K.2 applies this framework to the specific tests we conduct.

F Narrative Robustness: Granularity, Cross-Model, and Cross-Temperature

This appendix section reports the robustness of the narrative-extraction pipeline. First, we examine the stability of the results as we vary the granularity, i.e., the number of narratives requested. Second, we probe the consistency of the results across alternative LLM configurations.

Figure A6 presents the stability of the results across granularity levels. It shows the sunburst diagrams of the hierarchical structure of insertion and deletion narratives obtained by rerunning the pipeline with different requested numbers of narratives. The inner rings represent the broader classifications obtained when fewer narratives are requested, and the outer rings represent increasingly fine-grained classifications. The figure shows that all finer narratives remain nested within themes identified at coarser levels; none emerges as a new branch.

Figure A6: Hierarchical structure of narratives in insertions and deletions



Note: The figure shows narrative hierarchies for additions to Ruwiki, in green, and deletions from Wikipedia, in red. Inner rings correspond to broader narratives, while outer rings show finer-grained classifications produced by GPT-4.1 with temperature set to 0. Segment size reflects the model's confidence in the relationship between levels. Colors distinguish split and still-present narratives across levels; narratives without a coarser-level counterpart (*new*) would be highlighted by a bold black border, but none occurs in the baseline run. The stability of the hierarchy suggests that the extracted narratives reflect systematic patterns rather than arbitrary classifications.

We also re-run the pipeline using 13 robustness configurations. GPT-4.1, GPT-4o, and Claude Sonnet 4.5 are each run at temperatures 0, 0.5, and 1. OpenAI GPT-5 is run at low and high reasoning effort, OpenAI o3 at medium reasoning effort, and DeepSeek-V3 at temperature 1. All configurations, including the baseline, are accessed through a single API gateway (OpenRouter), which standardizes the request format across providers. Across these model variants, the pipeline produces a substantively similar set of narratives, differing primarily in wording. In each configuration, the cross-sectional analysis compares Russian-language Wikipedia articles with their Ruwiki counterparts to identify systematic content differences. For each direction (insertions and deletions), narratives are generated from the top 500 terms by χ^2 statistic, restricted to bigrams and trigrams; the model extracts seven narratives per direction and synthesizes an overall assessment across them.

Tables A2–A14 report the results, with each configuration identified in the corresponding table caption. Except for the English translations described below, all text in these tables is reproduced verbatim from the relevant model output and has not been edited by the authors.

Two configurations, GPT-5 with low reasoning effort and DeepSeek-V3, returned most of their output in Russian. As indicated in the corresponding table notes, the tables present English translations generated using GPT-4.1 at temperature 0. The original, unmodified Russian outputs are included in the replication package.

In several other cases, the models report the narratives in English but quote Russian input terms in their explanations. Each such term is followed by an English translation in square brackets, drawn from the study's term-translation table or, for terms not included in that vocabulary, generated using GPT-4.1 at temperature 0. Outside the two configurations whose outputs were translated in full, these bracketed English translations are the only text added by the authors to the tables.

Table A2: LLM Narrative Assessment (Robustness): GPT-4.1 with Temperature=0 (Independent Re-Run of the Baseline)

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The addition patterns reflect a strong push to formalize Russian administrative, legal, and demographic narratives in newly annexed or disputed territories, while reinforcing anti-LGBTQ+ and pro-sovereignty Russian legal frameworks. In contrast, deletions systematically remove references to Russian aggression, Ukrainian sovereignty, Western or opposition sources, LGBTQ+ content, and international legal or sanction-related information. This suggests a coordinated effort to align Wikipedia content with official Russian state narratives, minimize dissent, and erase Ukrainian and Western perspectives, especially in the context of the ongoing conflict and recent legislative changes.

Deletion Narratives	Insertion Narratives
1. Removal of References to Russian Invasion and Aggression: Deletions target explicit mentions of the Russian invasion of Ukraine, aggression, and related terminology, likely to reduce negative framing of Russian actions.	**1. Integration of New Russian Subjects:** Additions emphasize the formal inclusion of new territories (e.g., Donetsk, Luhansk, Kherson, Zaporizhzhia) into the Russian Federation, referencing legal status, administrative changes, and ongoing disputes with Ukraine.
2. Erasure of Ukrainian Sovereignty and Territorial Integrity: Content referencing Ukraine's territorial integrity, sovereignty, and international support for these positions is being removed, downplaying Ukraine's claims.	**2. Ongoing Territorial Disputes with Ukraine:** Content is added to highlight that Ukraine continues to dispute Russian claims over certain territories, with references to contested sovereignty and ongoing military actions.
3. Suppression of Western and Opposition Sources: References to independent or critical media, opposition figures, and Western organizations are deleted, reducing visibility of dissenting or critical perspectives.	**3. Legal and Administrative Reforms in Russian Regions:** Additions detail changes in administrative boundaries, legal statuses, and the organization of local governance, especially in regions like Chuvashia and newly annexed areas.
4. Censorship of LGBTQ+ and Sexual Content: Deletions remove mentions of LGBTQ+ identities, sexual orientation, and explicit sexual content, aligning with Russian censorship policies.	**4. Demographic and Census Data Updates:** New content provides updated demographic statistics, including gender breakdowns, population counts, and references to census data, often in the context of newly integrated or disputed territories.
5. Elimination of Ukrainian Administrative and Local Details: Details about Ukrainian administrative divisions, local governance, and place names are being deleted, erasing Ukrainian context from articles.	**5. Anti-LGBTQ+ Legislation and Propaganda:** Additions reflect the introduction or reinforcement of anti-LGBTQ+ laws, bans on 'propaganda', and restrictions on same-sex marriage and adoption, aligning with recent Russian legislative trends.
6. Removal of War and Conflict Impact Data: Deletions target information about the impact of war, such as casualties, displacement, and military actions, possibly to minimize perceived consequences.	**6. Formalization of Russian Legal and Constitutional Priorities:** Content is added to assert the primacy of Russian constitutional norms and legal frameworks, especially in the context of integrating new territories and overriding previous Ukrainian laws.
7. Omission of International Sanctions and Legal Proceedings: References to sanctions, international courts, and legal actions against Russia or Russian individuals are being removed, reducing awareness of international repercussions.	**7. International Context and Comparisons:** Additions situate Russian actions within a global context, referencing other countries, international organizations, and comparative legal or demographic data.

Notes: This configuration coincides with the baseline specification (GPT-4.1, temperature=0); the run is an independent replication of the baseline. Temperature 0 does not make the API deterministic across calls, so wording differs from Table 1 while the substance of the narratives replicates.

Table A3: LLM Narrative Assessment (Robustness): GPT-4.1 with Temperature=0.5

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The editing patterns in heavily edited Wikipedia articles reveal a strong push to align content with the official Russian narrative, especially regarding the incorporation of Ukrainian territories, legal and administrative changes, and the suppression of dissenting or Western-aligned perspectives. There is a notable effort to erase references to the Russian invasion, Ukrainian sovereignty, and LGBTQ topics, while simultaneously adding content that legitimizes Russian control, legal frameworks, and anti-LGBTQ measures. These patterns reflect a systematic information campaign aimed at consolidating Russian narratives and minimizing alternative or oppositional viewpoints.

Deletion Narratives	Insertion Narratives
1. Removal of References to Russian Invasion of Ukraine: Content directly mentioning the Russian invasion, war, or aggression in Ukraine is being systematically deleted, including references to the conflict's timeline and terminology such as 'вторжение россия украина [Russian invasion into Ukraine]'.	**1. Integration of New Russian Federal Subjects:** Content is being added to reflect the administrative and legal incorporation of new territories into the Russian Federation, especially regarding Donetsk, Luhansk, Kherson, and Zaporizhzhia regions, including the assignment of status, borders, and legal frameworks.
2. Erasure of Ukrainian Administrative and Legal Frameworks: Terms and codes specific to Ukrainian administration, such as KOATUU codes and Ukrainian local government structures, are being removed as Russian frameworks are imposed.	**2. Ongoing Territorial Disputes with Ukraine:** Additions highlight the contested nature of certain territories, emphasizing continued Ukrainian claims and disputes over sovereignty, as well as Russia's legal and administrative assertions.
3. Suppression of Western and Opposition Sources: References to media outlets critical of Russian policy, opposition figures, and Western organizations are being deleted, reducing mention of dissenting perspectives.	**3. Formalization of Local Administrative Units:** There is a focus on the creation, abolition, and reorganization of settlements, districts, and municipalities within Russia and annexed territories, including legal acts and status changes for local entities.
4. Reduction of LGBTQ and Sexuality-Related Content: Mentions of LGBTQ identities, sexual orientation, and related awards or scenes (especially in entertainment) are being deleted, aligning with anti-LGBTQ legislative trends.	**4. Legal and Constitutional Prioritization:** Content additions stress the precedence of Russian legal norms, constitutional provisions, and the implementation of Russian law in newly administered regions and broader federal subjects.
5. Elimination of Ukrainian Sovereignty and Integrity Assertions: References to Ukraine's territorial integrity, sovereignty, and international support for these positions are being removed, reflecting an effort to delegitimize Ukrainian claims.	**5. Demographic and Census Data Updates:** Additions provide updated population statistics, gender breakdowns, and census information for settlements, often reflecting new administrative boundaries or integrating newly acquired regions.
6. Removal of Historical and Demographic Contexts: Historical references (e.g., to wars, administrative history, or demographic data) that place current events in a broader or Ukrainian context are being deleted.	**6. Anti-LGBTQ and Propaganda Legislation:** Significant content is being added regarding laws and policies restricting LGBTQ rights, banning 'propaganda', and prohibiting same-sex marriage or adoption, aligning with recent legislative trends in Russia.
7. Deletion of Sanctions and International Legal Actions: Mentions of sanctions, international legal proceedings, and related punitive measures against Russia or Russian individuals are being removed, minimizing perceived external pressure.	**7. Geographical and Hydrological Detailing:** There is an influx of technical details about water resources, administrative codes, and geographical features, likely as part of standardizing and expanding local information in line with Russian administrative practices.

Notes: This robustness check uses GPT-4.1 with temperature=0.5 to assess sensitivity to the temperature parameter. Russian terms quoted in the model output are followed by an English translation in square brackets; the bracketed translations were added by the authors and are not part of the verbatim output.

Table A4: LLM Narrative Assessment (Robustness): GPT-4.1 with Temperature=1

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

Editing patterns reflect a re-framing of content towards legitimizing Russian sovereignty over disputed, recently annexed regions, while diminishing terms and perspectives highlighting invasion, Ukrainian sovereignty, civil society, LGBT rights, and alternative sources. There is systematic addition of local and legal terminology establishing Russian administrative structures and censorship of opposition, Western, and human rights language. The overall trajectory demonstrates a politicized edit war focused on aligning Wikipedia articles with Russian state narratives, while purging information and language supportive of Ukrainian sovereignty or critical of Russian actions.

Deletion Narratives	Insertion Narratives
1. Removal of References to Russian Invasion and Aggression: Deleted phrases explicitly mention 'invasion', 'aggression', 'annexation', and military operations by Russia against Ukraine, thereby diminishing the framing of conflict as an act of invasion.	**1. Legitimization of New Russian Territorial Entities:** Terms emphasize the formal inclusion of new territories and regions as 'subjects' of the Russian Federation, focusing on legal status, boundaries, and official recognition, especially concerning regions contested with Ukraine.
2. Suppression of Ukrainian Sovereignty and Integrity Language: Deleted terms frame Ukraine as sovereign and its territory as 'occupied' or violated, including references to Ukrainian integrity and related terms of international law and protest.	**2. Ukrainian Dispute and Russian Sovereignty Claims:** Added content asserts Russian claims over disputed areas, recurrently referencing Ukraine's ongoing contestation, and details scenarios where Russia 'continues to consider' these regions as its own despite international dispute.
3. Omission of Notable Anti-Kremlin and Western Media Figures/Sources: Content citing or referencing opposition figures, investigative journalists, and independent or Western media is removed, reducing visibility of critical or alternative perspectives.	**3. Integration of Donetsk, Luhansk, Kherson, and Zaporizhzhia:** Narratives detail specific legal and territorial changes for Donetsk (ДНР [DNR]), Luhansk (ЛНР [LNR]), Kherson, and Zaporizhzhia as parts of Russia, referencing related treaties, agreements, and local administrative structures.
4. Downplaying of Western or Ukrainian Recognition Systems: Deletions target standardized codes, administrative labels (e.g., KOATUU), Ukrainian local administrative details, and Western mapping/metadata frameworks, signaling a refocus on Russian-centric classification.	**4. Detailed Local Administrative and Legal Adjustments:** Content additions explain subdivisions, status changes, and legal acts concerning local settlements, municipal entities, new administrative units, and modifications to laws such as in Chuvash Republic.
5. Erasure of LGBT, Gender, and Sexuality Information: Removed content addresses LGBT issues, sexual orientation, and gender references, reflecting active filtering of queer representation and contemporary gender discourse.	**5. Denial and Prohibition of LGBT Rights:** Added passages include terms connected with prohibitions and negative stances toward LGBT rights, with references to anti-LGBT propaganda laws, bans, family law, and constitutional priorities.
6. Excision of Socio-legal Criticism and Human Rights References: Mentions of protests, human rights organizations, international courts, legal activism, and topics like human rights abuse or political repression are being systematically removed.	**6. Population, Census, and Demographic Structuring:** Several terms relate to documenting population figures, demographic compositions (e.g., by gender), and referencing official statistical systems for newly integrated or administratively altered territories.
7. Omission of Historical, Demographical and Civil Society Data: Removals include detailed demographical, community-based, historical, and civil society information, shifting focus away from the lived reality or history of local populations.	**7. Justification of Territorial Changes via International and Local Frameworks:** References are made to referenda, legal frameworks, and international comparisons, as well as Russia's rationale for territorial changes, including discussions of special military operations.

Notes: This robustness check uses GPT-4.1 with temperature=1 to assess sensitivity to the temperature parameter. Russian terms quoted in the model output are followed by an English translation in square brackets; the bracketed translations were added by the authors and are not part of the verbatim output.

Table A5: LLM Narrative Assessment (Robustness): GPT-4o with Temperature=0

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The patterns of additions and deletions in heavily edited Wikipedia articles reflect ongoing geopolitical tensions, particularly between Russia and Ukraine, with a focus on territorial changes and military actions. Additions emphasize new territorial incorporations and legal reforms, while deletions often involve sensitive or contested topics such as the Russian invasion of Ukraine. Cultural, demographic, and administrative changes are also prominent in additions, indicating shifts in governance and societal norms. Deletions frequently address outdated or redundant information, as well as personal and biographical details, reflecting a dynamic and evolving narrative landscape in these articles.

Deletion Narratives	**Insertion Narratives**
1. Russian-Ukrainian Conflict: Deletions focus on the Russian invasion of Ukraine and related military actions, possibly due to sensitivity or contested narratives.	**1. Territorial Changes in Russia and Ukraine:** Additions focus on the incorporation of new territories into the Russian Federation, particularly from Ukraine, including regions like Luhansk and Donetsk. This reflects ongoing geopolitical changes and disputes over territorial control.
2. Administrative and Geographic Data: Deletions include geographic and administrative data such as postal codes and local governance identifiers, possibly due to redundancy or updates.	**2. Legal and Administrative Reforms:** Content additions highlight legal and administrative changes within Russian territories, including the establishment of new administrative units and changes in territorial governance.
3. Cultural and Historical References: Deletions involve cultural and historical references, including awards and historical figures, possibly due to relevance or accuracy issues.	**3. Military Actions and Operations:** Additions include references to military actions, particularly in Ukraine, indicating ongoing conflicts and military operations affecting territorial status.
4. Personal and Biographical Information: Deletions include personal and biographical details, possibly due to privacy concerns or irrelevance to the main article topics.	**4. Cultural and Demographic Changes:** Additions reflect demographic and cultural shifts, including population statistics and cultural integration within newly incorporated regions.
5. Economic and Agricultural Data: Deletions focus on economic and agricultural data, possibly due to outdated information or shifts in economic focus.	**5. International Relations and Agreements:** Content additions discuss international agreements and diplomatic relations, particularly those affecting territorial claims and regional governance.
6. Media and Entertainment: Deletions involve media and entertainment references, such as film awards and reviews, possibly due to relevance or focus shifts.	**6. Local Governance and Infrastructure:** Additions focus on local governance structures and infrastructure development, including the establishment of new settlements and administrative boundaries.
7. Political and Legal Developments: Deletions include political and legal developments, possibly due to changes in political climate or legal status.	**7. LGBT Rights and Legislation:** Additions include discussions on LGBT rights and related legislation, reflecting societal debates and legal changes within Russia.

Notes: This robustness check uses GPT-4o with temperature=0 to assess sensitivity to the temperature parameter.

Table A6: LLM Narrative Assessment (Robustness): GPT-4o with Temperature=0.5

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The editing patterns in this category suggest a dynamic narrative surrounding the geopolitical and administrative changes involving Russia and Ukraine. Additions focus on expanding and legitimizing Russian territorial claims, legal structures, and cultural integration, while deletions aim to downplay controversial topics, historical invasions, and international criticism. This indicates a potential effort to shape public perception by emphasizing certain narratives while minimizing others, particularly those that might portray Russia in a negative light or highlight opposition to its actions. The content changes reflect a complex interplay of nationalistic, legal, and international themes.

Deletion Narratives	**Insertion Narratives**
1. Removal of Invasion References: Deletions focus on removing references to Russia's invasion of Ukraine, possibly to downplay or reframe the narrative.	**1. Expansion of Russian Federation:** Additions focus on the integration of new regions into the Russian Federation, including details about territories, legal status, and governance structures.
2. Sanitization of Controversial Topics: Content related to controversial topics such as LGBTQ issues and sexual orientation is being reduced or removed.	**2. Territorial Disputes with Ukraine:** Content additions highlight ongoing territorial disputes between Russia and Ukraine, emphasizing contested regions and claims of sovereignty.
3. Reduction of Local Administrative Data: Deletions include local administrative data such as postal codes and telephone indexes, possibly to streamline content.	**3. Formation of New Administrative Units:** Additions describe the administrative restructuring within Russia, including the creation and dissolution of settlements and regions.
4. Minimization of Historical Context: Historical references, particularly those related to past conflicts or governance, are being removed or minimized.	**4. Legal and Constitutional Changes:** This narrative involves the introduction of new laws and constitutional amendments affecting territorial governance and status.
5. Downplaying of International Criticism: Deletions reduce content related to international criticism or opposition to Russian actions, such as sanctions or protests.	**5. Military and Political Actions:** Additions cover military actions and political decisions impacting the territorial integrity and administrative control in contested regions.
6. Trimming of Cultural and Biographical Data: Biographical details and cultural references, particularly those not directly related to the main narrative, are being trimmed.	**6. Cultural and Demographic Information:** Content additions provide demographic data and cultural insights about the populations in newly integrated or disputed regions.
7. Editing of Sports and Entertainment Content: Content related to sports and entertainment, including player statistics and film awards, is being selectively removed.	**7. International Relations and Sanctions:** Additions reflect international responses to Russian territorial claims, including sanctions and diplomatic actions.

Notes: This robustness check uses GPT-4o with temperature=0.5 to assess sensitivity to the temperature parameter.

Table A7: LLM Narrative Assessment (Robustness): GPT-4o with Temperature=1

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The activity in this heavily edited category reflects significant geopolitical narratives involving the territorial disputes between Russia and Ukraine. Additions emphasize territorial and legal changes, Russian cultural influence, and demographic data, possibly to legitimize Russian claims. Deletions focus on censoring negative portrayals of Russian actions, downplaying territorial integrity violations, and removing references to contentious social issues. These patterns suggest an ongoing struggle over narrative control in the face of evolving geopolitical realities in Eastern Europe, with potential implications for both local governance and international diplomacy.

Deletion Narratives	Insertion Narratives
1. Removal of Russian Invasion Context: Deletions aim to downplay or remove the context regarding Russian military actions and invasions, particularly in Ukraine, by omitting specific references.	**1. Territorial Changes in Russia and Ukraine:** This narrative reflects the political dynamics around new administrative and territorial configurations involving Russia, Ukraine, and various peoples' republics. It highlights evolving geographic and political boundaries in response to territorial disputes or agreements.
2. Censorship of Sensitive Locations and Events: Content regarding sensitive political and military events, like the annexation of Crimea, is being diminished or erased.	**2. Integration and Dispute of Ukrainian Regions:** Content additions focus on contested areas such as Luhansk and Donetsk, emphasizing attempts to integrate these regions into Russian control while contesting Ukrainian sovereignty.
3. Reduction of Ukrainian Territorial Integrity Focus: Deletions minimize coverage on the importance of Ukrainian territorial integrity, indicating a potential narrative shift or censorship strategy.	**3. Legal and Governance Structures:** These additions highlight changes to legal frameworks, governance structures, and administrative designations within regions impacted by geopolitical shifts.
4. Obfuscation of Russian Criticism: Removal of content critical of Russian actions or leadership, including references to international sanctions or notable criticism.	**4. Military Actions and Operations:** Narrative on ongoing military engagements in Ukraine, indicating that the additions are possibly analyzing the impact of military action on territorial integrity and civilian life.
5. Elimination of Specific Organizational and Cultural References: Deletions of references to specific organizations or cultural details, perhaps to reduce international visibility or critique.	**5. Demographic and Census Data:** Focus on statistical analysis of population data, highlighting demographic changes possibly caused by political shifts or conflicts.
6. Reduction of LGBT Association Articles: Deletions aiming to remove or reduce visibility of content related to LGBT issues or individuals, possibly for sociopolitical reasons.	**6. Legal Status and Human Rights:** Attention to legal changes concerning human rights, possibly including same-sex relationships and family laws within Russian-influenced territories.
7. Simplification of Historical Context: Deletion of historical aspects related to Russia's geopolitical acts, aiming to simplify or omit their complex historical context.	**7. Russian Cultural and Political Influence:** Additions could reflect the influence of Russian cultural, legal, and political systems in newly integrated regions or disputed territories, highlighting integration attempts.

Notes: This robustness check uses GPT-4o with temperature=1 to assess sensitivity to the temperature parameter.

Table A8: LLM Narrative Assessment (Robustness): Claude Sonnet 4.5 with Temperature=0

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The editing patterns reveal a systematic effort to align Wikipedia content with Russian governmental narratives following the 2022 invasion of Ukraine. Additions extensively document Russia's annexation claims, administrative integration of occupied territories, and restrictive social legislation (particularly anti-LGBTQ+ laws), while framing military actions through official euphemisms. Deletions systematically remove terminology characterizing Russian actions as invasion or aggression, eliminate Ukrainian administrative information for occupied areas, and erase references to international condemnation and sanctions. The removal of opposition voices and independent media references suggests suppression of dissenting perspectives. The Chuvash Republic reorganization and Tatarstan cultural documentation additions appear unrelated to the Ukraine conflict but reflect ongoing Russian internal administrative changes. The contrast between technical, legalistic addition language and the removal of rights-based, sovereignty-affirming deletion terms indicates contested narratives about legitimacy, territorial control, and international law. These patterns suggest either coordinated pro-Russian editing or systematic removal of content deemed inconsistent with Russian governmental positions.

Deletion Narratives	Insertion Narratives
1. Russian Invasion Framing and War Terminology: Systematic removal of terms explicitly describing Russia's actions as 'invasion,' 'aggression,' or 'war' against Ukraine. Deletions eliminate language characterizing events as violations of Ukrainian sovereignty and territorial integrity. This suggests editorial pressure to adopt euphemistic Russian governmental terminology over international legal descriptions.	**1. Russian Annexation of Ukrainian Territories:** Extensive additions document Russia's incorporation of Ukrainian regions (DNR, LNR, Zaporizhzhia, Kherson) as new federal subjects. Terms emphasize legal frameworks, territorial definitions, and administrative structures while noting Ukraine's continued contestation of these territories. This reflects systematic documentation of Russia's territorial claims following the 2022 invasion.
2. Ukrainian Administrative and Contact Information: Removal of Ukrainian administrative codes (KOATUU), postal indices, telephone codes, and local council information for territories now under Russian control. This deletion erases Ukrainian governmental infrastructure documentation, replacing it with Russian administrative frameworks.	**2. Chuvash Republic Administrative Reorganization:** Significant content additions detail administrative boundary changes, municipal status modifications, and territorial reorganization within Chuvashia. Terms reference laws establishing new settlement boundaries, merging rural districts, and redefining urban-rural classifications. This suggests active restructuring of local governance units within this Russian republic.
3. Sanctions and International Condemnation: Deletions remove references to sanctions lists, international legal proceedings, and condemnation by Western institutions. Terms about European Court rulings, Canadian/US sanctions, and CIS Assembly positions are eliminated, obscuring international response to Russian actions.	**3. Anti-LGBTQ+ Legislation and Restrictions:** Additions extensively document Russia's expanding legal restrictions on LGBTQ+ communities, including propaganda laws, adoption bans, and constitutional amendments prioritizing traditional marriage. Terms indicate systematic codification of discrimination against same-sex relationships and gender identity expression, reflecting Russia's conservative social policy shift.
4. Opposition Figures and Critical Media: Removal of references to opposition politicians (Alexei Navalny), independent media outlets (Novaya Gazeta, Echo of Moscow, Radio Liberty), and critical voices. This deletion eliminates documentation of dissent and alternative perspectives on Russian government actions.	**4. Military Operations and Special Military Operation Context:** Content additions frame events within Russia's 'special military operation' in Ukraine, documenting military actions, territorial control changes, and wartime administrative adjustments. This narrative provides Russian governmental framing of the invasion as a legitimate operation rather than aggressive warfare.
5. Historical Ukrainian Village Demographics: Systematic deletion of detailed pre-Soviet and Soviet-era demographic data for Ukrainian villages, including household counts, estate classifications, economic activities, and ethnic compositions. This removes historical context that might establish Ukrainian cultural and administrative continuity.	**5. Water Resource Classification System:** Technical additions document Russia's water resource management framework, including basin codes, territorial classification systems, and federal agency oversight. Terms suggest standardization of hydrological data and administrative control over water resources, possibly related to infrastructure development or environmental monitoring.
6. Adult Entertainment Industry Content: Removal of detailed information about adult film industry awards (AVN), performer biographies, film titles, and career details. This suggests cleanup of explicit content or enforcement of content standards, possibly related to broader morality legislation.	**6. International Isolation and Comparative Legal Frameworks:** Additions reference Russia's alignment with countries having restrictive LGBTQ+ policies (Saudi Arabia, DPRK) and emphasize constitutional supremacy over international norms. This narrative positions Russia within a coalition of socially conservative states resisting Western liberal values.
7. LGBTQ+ Rights and Discrimination Documentation: Deletion of content documenting LGBTQ+ rights violations, discrimination cases, and comparative international perspectives on sexual orientation and gender identity. While additions emphasize restrictions, deletions remove human rights framing and international comparisons.	**7. Tatarstan Regional Identity and Academic Documentation:** Content additions emphasize Tatar cultural institutions, academic conferences, and regional historical documentation. Terms reference Kazan institutes, Tatar literature, and scholarly materials, suggesting efforts to document or promote Tatarstan's distinct identity within the Russian Federation.
8. Ukrainian Leadership and Sovereignty References: Removal of references to Ukrainian President Zelensky, Ukrainian government decisions, and assertions of Ukrainian sovereignty over contested territories. This deletion eliminates recognition of Ukrainian governmental authority in regions Russia claims.	

Notes: This robustness check uses Claude Sonnet 4.5 with temperature=0 to assess sensitivity to the temperature parameter.

Table A9: LLM Narrative Assessment (Robustness): Claude Sonnet 4.5 with Temperature=0.5

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The editing patterns reveal intense information warfare surrounding Russia's 2022 invasion and annexation of Ukrainian territories. Additions systematically document Russian administrative integration of occupied areas, anti-LGBTQ legislation, and legal frameworks asserting Russian sovereignty, while using euphemistic 'special military operation' terminology. Deletions remove explicit references to 'invasion', 'aggression', Ukrainian sovereignty, international sanctions, and opposition voices. This represents coordinated efforts to reshape Wikipedia's narrative: pro-Russian editors add legitimizing administrative details and legal justifications while removing critical terminology and Ukrainian governmental references. The Chuvash Republic restructuring and water resource coding appear as parallel bureaucratic documentation efforts. Adult content cleanup and historical data removal represent separate editorial concerns. Overall, the heavily-edited articles show Wikipedia as a contested space where geopolitical conflicts play out through systematic addition and removal of politically-charged terminology, with each side attempting to control the historical record and frame ongoing events according to their preferred narrative.

Deletion Narratives	Insertion Narratives
1. Russian Invasion Terminology Removal: Systematic removal of terms explicitly describing Russia's actions as 'invasion' or 'aggression'. Phrases like 'вторжение россия украина [Russian invasion into Ukraine]' (Russian invasion of Ukraine), 'российская агрессия [Russian aggression]' (Russian aggression), and 'война против украина [war against Ukraine]' (war against Ukraine) being deleted suggests replacement with neutral or pro-Russian framing. This represents editorial warfare over characterization of the conflict.	**1. Russian Annexation of Ukrainian Territories:** Terms reflect Russia's formal incorporation of Ukrainian territories (DNR, LNR, Kherson, Zaporizhzhia) as new federal subjects. References to treaties, territorial definitions, and Ukraine's disputed authority indicate documentation of annexation processes. Terms like 'новый субъект состав [new subject composition]' (new subject composition) and 'власть украина оспаривать [Ukrainian authority disputed]' (Ukraine disputes authority) show contested sovereignty being encoded into Wikipedia articles during Russia's 2022 territorial claims.
2. Ukrainian Administrative Information Purging: Removal of Ukrainian administrative codes, postal indices, and local council information from occupied territories. Terms like 'код коатуа [KOATUU code (Ukrainian administrative code)]', 'почтовый индекс [postal code]', and 'местный совет луганська область [Local council Luhansk region]' indicate deletion of Ukrainian governmental infrastructure references. This reflects replacement of Ukrainian administrative systems with Russian ones in articles about occupied areas.	**2. Administrative Restructuring in Chuvash Republic:** Extensive terms about Chuvash Republic's internal administrative reorganization, including settlement boundaries, status changes, and municipal formations. References to laws, territorial units, and settlement mergers indicate systematic documentation of local government restructuring. Terms like 'закон чувашский республика [law of the Chuvash Republic]' and 'граница сельский поселение [border rural settlement]' suggest detailed mapping of administrative changes within this Russian federal subject.
3. Territorial Integrity and Sovereignty References: Deletion of terms affirming Ukrainian territorial integrity and sovereignty. Phrases like 'целостность украина [integrity of Ukraine]' (Ukraine's integrity), 'защита целостность украина [integrity of Ukraine] [protection of Ukraines integrity]' (protection of Ukraine's integrity), and 'нарушение целостность [violation of integrity]' (violation of integrity) being removed suggests erasure of legal frameworks supporting Ukraine's borders. This aligns with Russian territorial claims.	**3. Anti-LGBTQ Legislation and Propaganda:** Terms document Russia's expanded anti-LGBTQ laws, including propaganda bans, adoption restrictions, and constitutional amendments. References to 'пропаганда лгбт [LGBT propaganda]', 'однополый брак [same-sex marriage]' (same-sex marriage), and 'усыновление ребёнок [child adoption]' (child adoption) indicate systematic addition of legal restrictions. International comparisons with countries banning LGBTQ rights suggest framing Russia's policies within global context of restrictive legislation.
4. Opposition Figures and Critical Media: Removal of references to Russian opposition figures and independent media outlets. Terms like 'алексей навальный [Alexei Navalny]', 'новый газета [Novaya Gazeta (Russian Opposition Newspaper)]', 'радио свобода [Radio Liberty (media outlet)]', and 'эхо москва [Echo Moscow (literal phrase, not the official radio station name)]' indicate deletion of content about regime critics and closed media. This suggests systematic removal of information about political opposition and independent journalism.	**4. Military Operations and Special Military Operation Context:** Terms reference Russia's 'special military operation' in Ukraine, providing context for territorial and administrative changes. Phrases like 'специальный военный операция [special military operation (official designation)]', 'военный действие украина [military actions involving Ukraine]', and 'фон военный действие [background of military operations]' indicate framing of events within official Russian terminology. This narrative contextualizes other changes as occurring during ongoing military operations.
5. Sanctions and International Condemnation: Deletion of information about international sanctions and legal consequences. Terms like 'санкционный список [sanctions list]', 'включить санкционный список [sanctions list] [add to the sanctions list (to include an entity in the list)]', and references to sanctions by USA, Canada, and EU indicate removal of documentation about international responses. This narrative shows erasure of consequences for Russia's actions.	**5. Water Resource Classification and Coding:** Technical terms about water resource management systems, basin numbering, and federal agency oversight. References to 'бассейн номер выпуск [basin number issue]', 'водный ресурс код [water resources code]', and 'федеральный агентство водный [Federal Water Agency]' indicate standardized documentation of hydrological data. This appears to be systematic addition of technical classification information to geographical articles.
6. Adult Entertainment Industry Content: Removal of detailed adult film industry information including AVN awards, performer biographies, and explicit content descriptions. Terms like 'премия [monetary prize (financial award)] avn', 'сайт [website (general term)] avn', 'анальный секс [anal sex]', and 'лесбийский сцена [lesbian-themed scene (scene with lesbian content or focus)]' indicate cleanup of pornographic content details. This suggests enforcement of content standards or removal of promotional material from performer articles.	**6. Constitutional Primacy and Legal Framework:** Terms emphasize Russian constitutional supremacy over international norms and family law provisions. References to 'приоритет российский конституция [priority of the Russian Constitution]', 'семейный кодекс [Family Code]', and 'положение приоритет [priority position]' suggest documentation of legal hierarchy placing Russian law above international obligations. This narrative reinforces domestic legal authority in contested areas.

(continued on next page)

(Table A9 continued from previous page)

Deletion Narratives	Insertion Narratives
7. Historical Administrative and Economic Data: Removal of detailed historical information about pre-Soviet and Soviet-era administrative divisions, economic activities, and settlement structures. Terms like 'уезд казанский губернии [uyezd of Kazan Governorate]', 'кантон тасср [canton TASSR (Tatar Autonomous Soviet Socialist Republic)]', 'сельский община [village commune (traditional rural self-governing community)]', and 'личный подсобный хозяйство [personal subsidiary farm]' indicate deletion of granular historical data. This may reflect standardization efforts or removal of outdated information.	**7. Foreign Agent Registry and Restrictions:** Terms document Russia's foreign agent legislation and restrictions on organizations and individuals. References to 'реестр иностранный агент [register of foreign agents (legal designation)]', 'движение запрещать рф [movement banned in the Russian Federation]', and 'минюст рф [Ministry of Justice of the Russian Federation (official state body)]' indicate systematic labeling and legal restrictions. This narrative shows documentation of government controls over civil society and media organizations deemed foreign-influenced.

Notes: This robustness check uses Claude Sonnet 4.5 with temperature=0.5 to assess sensitivity to the temperature parameter. Russian terms quoted in the model output are followed by an English translation in square brackets; the bracketed translations were added by the authors and are not part of the verbatim output.

Table A10: LLM Narrative Assessment (Robustness): Claude Sonnet 4.5 with Temperature=1

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The editing patterns reveal a systematic transformation of Wikipedia content related to Russian-occupied Ukrainian territories and domestic Russian policies. Additions extensively document Russian administrative claims over annexed regions using formal bureaucratic language while emphasizing legal frameworks that present territorial incorporation as legitimate. Simultaneously, deletions systematically remove Ukrainian administrative identifiers, explicit invasion/war terminology, and international legal responses.

A parallel narrative emerges around LGBTQ+ content: additions extensively document restrictive Russian legislation while deletions remove contextual information about rights and discrimination. The patterns suggest coordinated efforts to: (1) normalize territorial annexation through administrative documentation, (2) control conflict terminology by removing ”invasion” language, (3) erase Ukrainian administrative legitimacy, (4) minimize international legal consequences, and (5) reshape LGBTQ+ content to emphasize restrictions while removing rights-based framing.

Additional patterns include removal of opposition figures and independent media references, extensive deletion of adult entertainment content, and streamlining of historical details. These coordinated editing patterns across multiple topic areas suggest systematic narrative management rather than organic editorial evolution.

Deletion Narratives	**Insertion Narratives**
1. Russian Invasion Terminology Removal: Systematic deletion of terms explicitly describing Russian military actions as ”invasion,” ”aggression,” or ”war.” Removal of phrases linking Russia to attacks on Ukraine suggests efforts to sanitize language around military operations, replacing direct terminology with euphemistic or administrative language. Reflects narrative control around characterization of the conflict.	**1. Russian Annexation of Ukrainian Territories:** Extensive additions of formal bureaucratic language documenting Russia’s incorporation of Ukrainian regions (DNR, LNR, Zaporizhzhia, Kherson) into the Russian Federation. Terms emphasize legal frameworks, administrative boundaries, and territorial definitions within Russian federal structure, reflecting systematic documentation of annexed territories as new Russian federal subjects despite Ukrainian and international disputes.
2. Ukrainian Administrative Information Erasure: Deletion of Ukrainian administrative data including postal codes, telephone codes, KOATUU codes, local councils, and oblast references. Removal of these identifiers suggests systematic erasure of Ukrainian administrative frameworks from articles about occupied territories, replacing Ukrainian systems with Russian equivalents.	**2. Administrative Restructuring in Chuvash Republic:** Detailed additions about municipal reorganization, border changes, and status modifications of settlements in Chuvashia. Language focuses on legal frameworks establishing, abolishing, or modifying administrative units, rural and urban settlements, and municipal boundaries through regional legislation. Reflects systematic documentation of internal Russian administrative reforms.
3. Territorial Integrity and International Response: Removal of language about Ukrainian territorial integrity, international sanctions against Russia, and global responses to annexation. Deletions include references to sanctions lists, UN resolutions, European court actions, and protection of Ukraine’s sovereignty, suggesting minimization of international legal and political consequences.	**3. Military Operations Context:** Additions framing content within the context of Russia’s military operations in Ukraine, using euphemistic language like ”special military operation.” Terms establish temporal and contextual frameworks for territorial changes, suggesting systematic efforts to normalize military actions through bureaucratic documentation and neutral-sounding administrative language.
4. Russian Dissident and Opposition References: Deletion of mentions of opposition figures like Alexei Navalny, independent media outlets (Novaya Gazeta, Echo of Moscow, Radio Liberty), and criminal cases. Removal suggests systematic cleansing of content related to political opposition, independent journalism, and government critics.	**4. Anti-LGBTQ+ Legal Framework:** Extensive documentation of Russian legislation restricting LGBTQ+ rights, including propaganda laws, adoption prohibitions, and recognition of LGBTQ+ movements as extremist. Language emphasizes legal prohibitions, family code restrictions, and constitutional prioritization of traditional values, reflecting systematic legal marginalization of LGBTQ+ communities.
5. Adult Entertainment Industry Content: Extensive removal of adult film industry terminology including AVN awards, performer biographies, film scenes, and career details. Deletions encompass professional terminology, industry recognition, and biographical information, suggesting content moderation or editorial policy changes regarding adult entertainment coverage.	**5. Hydrological Classification Systems:** Technical additions of water resource coding systems, basin classifications, and federal agency data. Represents bureaucratic standardization of geographical and environmental information through systematic classification schemes, possibly related to resource management or territorial documentation projects.
6. Historical Administrative and Economic Details: Removal of historical details about pre-Soviet and early Soviet administrative divisions, economic structures, and settlement classifications. Deletions include references to uezds, volosts, TASSR cantons, peasant estates, and land measurements, suggesting streamlining of historical administrative detail.	**6. Population and Demographic Documentation:** Additions of detailed demographic data including gender breakdowns, census information, and population statistics for settlements. Language emphasizes permanent residents, male/female ratios, and village populations, suggesting systematic demographic documentation possibly related to administrative claims or census standardization.
7. LGBTQ+ Rights and Discrimination Context: Deletion of terms describing LGBTQ+ rights, sexual orientation, discrimination, and violence against LGBTQ+ individuals. Removal includes references to same-sex couples, lesbian and gay communities, and rights violations, possibly reflecting editorial changes to minimize discussion of LGBTQ+ issues or discrimination.	**7. International Relations and Sanctions:** Documentation of Russia’s international standing, including lists of countries with similar LGBTQ+ policies, foreign agent registries, and geopolitical alignments. Terms reference Central Asia, Africa, North Korea, and Saudi Arabia, suggesting efforts to contextualize Russian policies within selective international comparisons and emphasize alternative alliances.

Notes: This robustness check uses Claude Sonnet 4.5 with temperature=1 (default) to assess sensitivity to temperature parameter.

Table A11: LLM Narrative Assessment (Robustness): OpenAI GPT-5 with Reasoning Effort=Low

Except for the English translations described in the table note, all text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

Additions concentrate on the legal integration of disputed territories into the legal field of the Russian Federation, duplicating formulas about composition, borders, and the priority of Russian legislation, while simultaneously recording Ukrainian disputes. In parallel, municipal reforms (especially in Chuvashia), demographic blocks, and technical-reference standardization are strengthened; a shift toward conservative socio-legal norms (anti-LGBT, foreign agents) is noticeable. Deletions remove explicit military and sanctions framing, sections on human rights and opposition, pornographic content and sports chronicle, as well as Ukrainian administrative codes and universal references. Overall, there is an institutionalization of the new map and regulations while simultaneously cleansing conflict-political and 'inappropriate' details.

Deletion Narratives	Insertion Narratives
1. Removal of formulations about 'Russia's invasion of Ukraine': Words such as 'invasion', 'Russian-Ukrainian war', 'for the invasion', mentions of the annexation of Crimea and the course of hostilities are systematically removed. Conflict interpretation and political-evaluative formulas, including sanctions and responsibility, are reduced.	**1. Integration of new subjects of the Russian Federation and legalization of annexations:** Language about 'new subjects' and the inclusion of the DPR, LPR, Kherson and Zaporizhzhia regions into the Russian Federation is added. Formulas such as 'the Russian Federation determines the territory/region', references to 'a decision based on the results of referendums', and administrative linkage to Russian norms and constitutional priority are clarified.
2. Removal of Ukrainian administrative codes and reference attributes: Postal codes, telephone codes, KOATUU, and links to Ukrainian councils/communities are removed. This weakens Ukrainian administrative classification and reference blocks in articles about settlements.	**2. Ukraine disputes sovereignty and retains its borders de jure:** In parallel, Ukraine's position is fixed: 'disputes', 'continues to consider as territory/region of Ukraine', including the DPR/LPR. Dual status and ongoing military actions are recorded, showing the conflictual and unrecognized nature of the changes.
3. Cleansing of sports 'game chronicle': Minor details about matches are removed: starting lineups, minute-by-minute substitutions, tournament statistics, final placements. Excessive chronicle and trivial facts of athletes' biographies are reduced.	**3. Municipal reorganization and territorial structure (Chuvashia and others):** Mass edits about the statuses and borders of settlements: abolition/creation, merging into urban okrugs, establishment of borders and powers. Emphasis on regional legislation of the Chuvash Republic and units of local self-government.
4. Exclusion of content about the adult industry and sexual scenes: Mentions of AVN, 'good scene', 'anal sex scene', 'adult film', 'new starlet', as well as details about studios and awards are removed. The explicitness and sexualization of articles is reduced.	**4. Detailing the composition of settlements and administrative affiliation:** Formulas such as 'region including village/settlement/urban-type settlement', 'part of the settlement/district/okrug' are added, subordination to 'city of significance' is clarified, and list attributes for settlements and districts, including in new subjects, are specified.
5. Removal of blocks about human rights, sanctions, and opposition: References to sanctions lists (USA/Canada/EU), human rights topics, 'freedom of speech', Navalny, ECHR court cases are removed. The political-legal context and international reaction are reduced.	**5. Demographic and census data:** Population indicators are clarified: 'permanent resident', 'composed of men/women', 'resident per capita', references to censuses. These additions standardize the statistical blocks of articles about settlements.
6. Reduction of historical-estate and agrarian details: Archaic attributes are removed: 'state village', 'stanovy apartment', 'desyatina', volosts/uyezds, manor-farm development. Historical-economic descriptions of settlements are simplified.	**6. Conservative social policy and prohibitions:** Sections on Russian legislation are expanded: ban on 'LGBT propaganda', ban on same-sex marriage and adoption, 'foreign agents', fight against drugs and 'feelings of believers'. Departmental and legal formulas are recorded.
7. Removal of universal references and climate statistics: 'Population census', 'average air temperature', transport accessibility of villages, telephone/post are removed. The probable goal is to combat templated, outdated, or irrelevant content.	**7. Technical regulations and reference data:** Official codes and registries (water resources, basins, GI), 'official code', 'federal agency for water resources', as well as lists of TV series episodes and dubbing information are added. Articles are formatted as reference guides.

Notes: This robustness check uses OpenAI GPT-5 with reasoning_effort=low for cross-model validation. This configuration returned part or all of its output in Russian; those passages are reported in the authors' English translation (GPT-4.1, temperature 0). The original Russian output is preserved unmodified in the replication package.

Table A12: LLM Narrative Assessment (Robustness): OpenAI GPT-5 with Reasoning Effort=High

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

Additions center on legal-administrative consolidation under Russian frameworks: asserting new 'subjects,' detailing DNR/LNR/Kherson/Zaporizhzhia down to settlement level, and documenting municipal reforms (notably in Chuvash). They also expand technical codifications (water-basin codes) and structured reference content (episodes, demographics). Deletions reverse or sidestep Ukrainian-centric and conflict-culpability framing, remove Ukrainian admin identifiers, and prune sanctions/rights/opposition-media mentions. They also strip adult/LGBT content, sports minutiae, archaic administrative notes, and local trivia. Overall, heavily edited pages are being reshaped toward Russian legal standards, standardized datasets, and tighter content scope, while de-emphasizing politically sensitive framings and granular or promotional details.

Deletion Narratives	Insertion Narratives
1. Removal of 'invasion/aggression' framing of the Ukraine war: Edits delete explicit labels of Russia's actions as an 'invasion' or 'aggression,' and references to Ukraine's territorial integrity and Crimea's annexation. Conflict sections lose phrasing tied to culpability and advocacy (e.g., 'support for the invasion').	**1. Legal framing of new Russian 'subjects' and territorial claims:** Additions stress incorporation of territories as new subjects of the Russian Federation, using legalistic phrasing that Russia determines borders and constitutional primacy applies. Parallel notes state Ukraine disputes these moves and continues to consider the areas its territory. References to September decisions and referendum outcomes reinforce the legal-justificatory frame.
2. Stripping Ukrainian administrative and infobox data: Ukrainian-specific datasets are removed: postal/telephone codes, KOATUU, and local councils/communes. This aligns with replacing Ukrainian administrative frameworks and identifiers in locality articles.	**2. Granular delineation of DNR/LNR/Kherson/Zaporizhzhia:** Extensive detailing of Donetsk and Luhansk 'people's republics' and Kherson/Zaporizhzhia areas: specifying territories, linkage to oblasts, and listing included settlements (villages, towns, urban-type). These entries codify composition and boundaries while reiterating Russian status assignments.
3. Pruning sanctions, rights, and opposition-media framing: Mentions of sanction lists, human-rights/legal institutions, and independent/opposition media are cut. This reduces content about external accountability, legal judgments, or political critique.	**3. Chuvash Republic municipal status and boundary reforms:** Many additions cite Chuvash Republic laws that establish, abolish, or redefine municipal entities and their borders, and assign urban/rural status. This standardizes local governance structures, often via legal formulas and boundary descriptions.
4. Scrubbing sexual/AVN/LGBT material: Biographical and filmography content about adult entertainment (AVN awards, explicit scene types) and LGBT rights/identities is removed, likely reflecting stricter sourcing and legal-compliance pressures.	**4. Nationwide consolidation into urban okrugs:** Articles gain material on municipal reorganization: merging settlements and districts into urban okrugs, defining their scope "within district frameworks," and clarifying inclusion of all settlements. Emphasis is on administrative consolidation and streamlined local governance units.
5. Trimming match□by□match sports minutiae: Fine-grained game logs (starting lineups, substitutions by minute, individual assists) are deleted in favor of more concise summaries, reducing excessive detail in athlete/club articles.	**5. Expansion of anti□LGBT restrictions and gender policies:** Content adds references to bans on LGBT 'propaganda,' adoption by LGBT people, and restrictions on gender transition, aligning with recent Russian federal laws. Entries also mark movements as prohibited and adjust biographies/policies accordingly.
6. Dropping imperial-era and archaic administrative details: Historical units and measures (uezd/volost, desyatina, kazyonny village, 'dvor,' stanovy quarters) are removed from locality pages, likely as undue or anachronistic for modern descriptions.	**6. Standardization of hydrological and territorial coding:** Technical additions introduce federal water-resource classifications and basin codes, citing releases and coding schemas by the Federal Agency for Water Resources. This enriches articles with official identifiers and dataset linkages.
7. Cutting locality trivia and generic climate/infrastructure: Generic or non-notable details (climate templates, average temperatures, bus stops, houses of culture) are pared back to streamline small-settlement articles and avoid unsourced trivia.	**7. Structured media and demographic data expansion:** Articles increasingly feature structured lists and metadata: TV episode lists and dubbing credits, plus demographic updates with gendered population breakdowns drawn from censuses. The trend emphasizes standardized, list-like reference content.

Notes: This robustness check uses OpenAI GPT-5 with reasoning_effort=high for cross-model validation.

Table A13: LLM Narrative Assessment (Robustness): OpenAI o3 with Reasoning Effort=Medium

All text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

Additions concentrate on recasting territory—formally integrating occupied Ukrainian regions, redrawing Russian municipal borders and inserting anti-LGBT legal language—while embedding supportive demographic and diplomatic data. Deletions mirror this by eliminating wording that frames Russia as an invader, scrubbing sanction references and removing LGBT or sexually explicit material, along with granular Ukrainian administrative or football trivia. Together, the edit patterns suggest an effort to align Wikipedia content with official Russian state narratives, minimise politically damaging material, and tidy peripheral details that conflict with or distract from that framing.

Deletion Narratives	Insertion Narratives
1. Erasure of 'Invasion/Aggression' Language: Words that label Russia's actions as an invasion, aggression or annexation are systematically stripped, softening descriptions of the conflict with Ukraine.	**1. Incorporation of Occupied Ukrainian Regions:** Large blocks of text describe LNR, DNR, Kherson and Zaporizhzhia as "new subjects" whose borders, treaties and status are now defined by the Russian Federation, while noting that Ukraine continues to dispute this claim.
2. Removal of Sanctions and Zelensky Decrees: Lists of Western sanctions, presidential orders and related legal details concerning Russia's international isolation are being cut from articles.	**2. Chuvash Republic Border & Status Laws:** Many edits add regional statutes that fix borders, grant or revoke urban/rural status and reorganise settlements inside the Chuvash Republic, reflecting an ongoing legal tidy-up of its internal territorial structure.
3. Purging Pornography Awards & Explicit Scenes: Entire sections on AVN honours, explicit sexual acts and porn-industry credits are deleted, reducing sexual detail and adult-film promotion.	**3. Municipal Mergers into Urban Okrugs:** Content details the dissolution and amalgamation of village councils and districts into larger 'городской [urban]/муниципальный округ [municipal district (general term)]', listing which settlements are merged and the legal acts that enforce it.
4. Deletion of LGBT Orientation & Rights Content: Mentions of gay, lesbian or bisexual people, orientation data and LGBT rights references are removed, mirroring the new anti-propaganda emphasis in additions.	**4. Emphasis on Anti-LGBT Norms:** New passages cite constitutional primacy, Family Code clauses and bans on 'ЛГБТ-пропаганда [LGBT propaganda]', same-sex marriage and adoption, positioning these prohibitions as federal legal priorities.
5. Stripping Technical Administrative Codes: Postal indexes, telephone codes and Ukrainian KOATUU identifiers are cut, possibly seen as trivia or politically sensitive Ukrainian administrative data.	**5. Expansion of Statistical & Registry Data:** Editors append technical registries—water-basin codes, census gender tallies, official settlement IDs—aimed at standardising factual datasets across articles.
6. Trimming Detailed Sports Match Statistics: Fine-grained football data—starting elevens, substitution minutes, match-by-match notes—is excised to streamline athlete biographies and club pages.	**6. Showcasing Global-South Diplomacy:** Additions highlight Russia's outreach to Central Asia, Africa, Saudi Arabia and the DPRK, stressing cooperative forums and diplomatic roles to frame Russia as globally engaged despite Western isolation.
7. Dropping Rural Zoning & Subsistence Farming Info: Descriptions of personal subsidiary farming, usadba zoning and local agrarian economies are removed, perhaps deemed unnecessary minutiae for general readers.	**7. Justifying the "Special Military Operation":** Text references the start of hostilities, frames them as a 'special operation', and notes ongoing military actions, aligning article language with official Russian terminology.

Notes: This robustness check uses OpenAI o3 with reasoning_effort=medium for cross-model validation. Russian terms quoted in the model output are followed by an English translation in square brackets; the bracketed translations were added by the authors and are not part of the verbatim output.

Table A14: LLM Narrative Assessment (Robustness): DeepSeek-V3 with Temperature=1

Except for the English translations described in the table note, all text in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The patterns of additions and deletions in the 'all_heavily_edited' category reflect a complex interplay of narratives related to Russian territorial claims, internal administrative reorganization, military actions in Ukraine, and legal changes. Additions tend to emphasize Russian administrative and territorial narratives, while deletions often remove content related to the Russian invasion of Ukraine, political opposition, LGBTQ+ rights, and local governance disputes. This suggests a strong editorial focus on controlling narrative around sensitive geopolitical issues and internal policies, while removing content that could be politically contentious or critical of Russian actions both domestically and internationally.

Deletion Narratives	Insertion Narratives
1. Russian invasion of Ukraine: Frequent deletions of terms like 'вторжение россия [Russian invasion]', 'российско украинский война [Russo-Ukrainian War]', and 'российский вторжение [Russian-led invasion]' suggest a narrative of removing sensitive content related to the invasion.	**1. Russian territorial claims and annexations:** The frequent mention of terms like 'ЛНР [LNR]', 'ДНР [DNR]', 'Херсонская область [Kherson Region]', and 'ЗАПОРОЖСКАЯ ОБЛАСТЬ [Zaporozhye Region]' suggests a narrative of Russian claims and administrative changes in disputed Ukrainian territories.
2. Sanctions and political opposition: Deletions of 'алексей навальный [Alexei Navalny]', 'санкционный список [sanctions list]', and 'эхо москва [Echo Moscow (literal phrase, not the official radio station name)]' indicate narratives around removing content related to political opposition and sanctions.	**2. Internal administrative reorganization:** Terms such as 'новый субъект [new subject]', 'состав российский федерация [part Russian Federation]', and 'область состав [region composition]' indicate a focus on internal restructuring within Russian administrative regions.
3. Historical conflicts and annexations: Terms like 'аннексия крым [annexation Crimea]', 'Великий война [Great War]', and 'российский оккупация [Russian occupation]' highlight deletions related to historical conflicts and territorial annexations.	**3. Legal and constitutional changes:** References to 'Russian constitution', 'Chuvash law', and 'federation determines' highlight narratives around legal and constitutional amendments.
4. LGBTQ+ rights and recognition: Deletions of 'сексуальный ориентация [sexual orientation (general term for sexual preference)]', 'право лгбт [LGBT right]', and 'левбиянка гей [Lesbian gay]' suggest narratives around removing content related to LGBTQ+ rights.	**4. Military actions in Ukraine:** Terms like 'военный действие украина [military actions involving Ukraine]' and 'специальный военный операция [special military operation (official designation)]' underscore narratives related to Russian military operations in Ukraine.
5. Local governance and community disputes: Deletions of 'сельский община [village commune (traditional rural self-governing community)]', 'местный совет [local council]', and 'волость уезд [volost uyezd (administrative divisions in Imperial Russia)]' indicate narratives around removing content related to local governance disputes.	**5. Local governance and community structure:** Frequent mentions of 'сельский поселение [village settlement (administrative unit)]', 'граница сельский [rural border]', and 'местный граница [local border]' suggest narratives focused on local governance and community organization.
6. Economic and agricultural policies: Terms like 'экономика сельский [rural economy (economic system of rural areas)]', 'сельский округа [rural districts (plural, administrative)]', and 'сельский совет [village council (local rural governing body)]' highlight deletions related to economic and agricultural policies.	**6. Historical and cultural narratives:** Terms like 'Tatar Khasanov', 'Tatar literature', and 'culture house of culture' indicate a focus on historical and cultural narratives within the region.
7. Cultural and historical deletions: Deletions of 'история статус [history status]', 'культура дом культура [House of Culture (Soviet/Russian institution for community arts and education)]', and 'наследие пункт [Heritage point]' suggest narratives around removing cultural and historical content.	**7. Economic and resource management:** Mentions of 'ресурс код [resource code]', 'водный ресурс [water resource]', and 'федеральный агентство водный [Federal Water Agency]' highlight narratives around economic policies and resource management.

Notes: This robustness check uses DeepSeek-V3 (via OpenRouter) with temperature=1 for cross-model validation. This configuration returned part or all of its output in Russian; those passages are reported in the authors' English translation (GPT-4.1, temperature 0). The original Russian output is preserved unmodified in the replication package. Russian terms quoted in the model output are followed by an English translation in square brackets; the bracketed translations were added by the authors and are not part of the verbatim output.

G Narrative Placebo: Wikipedia changes over time

This appendix section reports the placebo narrative-extraction results for changes within Wikipedia between the fork on May 1, 2023, and January 1, 2025 (Table A15). The analysis uses the same sample of 56,835 deleted and heavily edited articles identified in the January 2025 cross-sectional comparison. We also show that the narratives emerging from changes within Ruwiki over the same period in the same set of articles closely resemble those identified in the baseline Ruwiki–Wikipedia comparison (see Table A16 for within-Ruwiki comparison versus the baseline presented in Table 1).

G.1 Applying the Conceptual Framework to the LLM Analysis

When we apply the narrative-difference pipeline to the two comparisons (baseline vs. placebo), the interpretation in light of the conceptual framework developed in Online Appendix Section E is straightforward. We find substantial narrative differences in the cross-platform comparison but not in the within-Wikipedia comparison. This pattern indicates that the narrative differences separating Ruwiki from Wikipedia are not simply the result of ordinary temporal change in Wikipedia between the fork and January 2025. Instead, they point to active divergence by Ruwiki.

Table A15: Placebo LLM Narrative Assessment: Wikipedia over Time, January 2025 vs. the Fork

All text below is reproduced verbatim from GPT-4.1 (temperature 0) and has not been edited by the authors.

Overall Assessment

The editing patterns in this category reflect a dual movement: additions are expanding articles with rich historical, demographic, cultural, and sports-related content, often emphasizing local detail, recent geopolitical events, and recognition of achievements. In contrast, deletions focus on streamlining articles by removing redundant, overly technical, or bureaucratic information, as well as trimming less notable or duplicative content. This suggests a community effort to balance comprehensiveness with clarity and relevance, ensuring articles are both informative and accessible while maintaining notability and neutrality, especially regarding sensitive political topics.

Deletion Narratives

1. Removal of Outdated or Redundant Administrative Data: Deletions target local council information, administrative codes, and addresses, likely to eliminate outdated, redundant, or overly bureaucratic details from articles.

2. Pruning of Gendered and Redundant Demographic Statements: Content specifying gender breakdowns and repetitive population statements is being removed, possibly to streamline demographic sections and avoid unnecessary detail.

3. Reduction of Official and Bureaucratic References: References to federal agencies, official codes, and resource management are being deleted, suggesting a move away from technical or bureaucratic language in favor of more accessible content.

4. Elimination of Redundant or Duplicative Population Data: Multiple deletions involve repetitive or duplicative statements about population numbers, likely to improve clarity and avoid confusion in articles.

5. Trimming of Minor or Peripheral Biographical Details: Details about secondary roles, dubbing, and minor awards are being removed, indicating a focus on streamlining biographical entries to emphasize primary achievements.

6. Removal of Peripheral or Non-Notable Localities: References to small or declining settlements, gradual population decreases, and minor administrative units are being deleted, possibly to maintain notability standards.

7. Reduction of Political and Conflict-Related Content: Some deletions target references to sanctions, political conflicts, and controversial figures, suggesting editorial caution or attempts to depoliticize certain articles.

Insertion Narratives

1. Expansion of Rural and Agricultural Histories: Many additions focus on rural settlements, agricultural practices, and administrative divisions, reflecting a drive to document local histories, economic bases, and demographic details of villages and rural communities.

2. Detailed Sports and Athlete Coverage: There is a significant increase in content about sports events, athletes' careers, match details, and achievements, indicating a trend toward comprehensive sports documentation and recognition of individual and team accomplishments.

3. Emphasis on War and Geopolitical Events: Additions highlight recent and historical conflicts, especially the Russian invasion of Ukraine, World Wars, and related sanctions, suggesting a focus on updating articles with current events and their historical context.

4. Recognition of Achievements and Honors: There is a notable increase in mentions of awards, honors, and distinctions, reflecting a narrative of celebrating individual and collective achievements in various fields, including military, sports, and culture.

5. Demographic and Census Data Enrichment: Additions frequently include detailed population statistics, census results, and ethnic/language breakdowns, indicating a push for more granular and up-to-date demographic information in articles.

6. Cultural and Religious Heritage Documentation: There is an increased focus on cultural institutions, religious sites, and heritage, suggesting a narrative of preserving and highlighting local traditions, religious practices, and notable landmarks.

7. Integration of Critical Reception and Media Analysis: Additions include critical reviews, film ratings, and media-related content, reflecting a trend toward documenting public and critical reception of cultural products and notable media events.

Notes: This table reproduces verbatim the output of the GPT-4.1 narrative-extraction pipeline with temperature set to 0. The model's original wording, narrative labels, and overall assessment have not been edited and should not be interpreted as author-written exposition. The within-Wikipedia placebo examines changes in Wikipedia between the fork date (May 1, 2023) and January 1, 2025. The placebo uses the same set of 56,835 articles as the cross-sectional analysis, tracking their within-Wikipedia changes. The LLM identified seven deletion narratives and seven insertion narratives from the 500 terms with the highest χ^2 statistics in each direction, using bigrams and trigrams only.

Table A16: LLM Narrative Assessment: Ruwiki over Time, January 2025 vs. the Fork

All text below is reproduced verbatim from GPT-4.1 (temperature 0) and has not been edited by the authors.

Overall Assessment

The addition patterns reflect a strong push to legitimize Russian control over disputed territories, reinforce Russian legal and administrative frameworks, and promote state-sanctioned narratives, including anti-LGBT and pro-Russian identity themes. Deletions systematically erase Ukrainian administrative presence, suppress narratives of Ukrainian sovereignty, resistance, and democracy, and remove references to independent media, human rights, and international legal scrutiny. This editing dynamic reveals a concerted effort to reshape Wikipedia content in line with official Russian state narratives, while actively suppressing alternative, pro-Ukrainian, and liberal-democratic perspectives.

Deletion Narratives

1. Removal of Ukrainian Administrative Data: Deletions target Ukrainian administrative identifiers such as postal codes, KOATUU codes, and local council references, erasing traces of Ukrainian governance in contested regions.

2. Erasure of Ukrainian Sovereignty and War Framing: Content referencing the Russian invasion, Ukrainian territorial integrity, and the framing of the conflict as aggression or occupation is systematically removed.

3. Suppression of Pro-Ukrainian and Democratic Narratives: Mentions of Ukrainian resistance, democratic values, and civil society (e.g., protests, Revolution of Dignity, rights activism) are deleted to minimize narratives of Ukrainian agency.

4. Censorship of Independent and Opposition Media: References to independent Russian and Ukrainian media outlets, as well as opposition figures and organizations, are removed to limit alternative viewpoints.

5. Elimination of LGBT and Human Rights Discourse: Deletions focus on removing mentions of LGBT rights, sexual orientation, and related organizations, as well as broader human rights and anti-discrimination language.

6. Obfuscation of Population and Social Statistics: Population data, census results, and demographic information are deleted, possibly to obscure the impact of conflict or administrative changes on local populations.

7. Removal of Legal and International Accountability References: Mentions of international law, courts, sanctions, and legal proceedings against Russia or its officials are deleted to avoid narratives of accountability or illegitimacy.

Insertion Narratives

1. Integration of New Russian Territories: Additions emphasize the formal inclusion of Donetsk, Luhansk, Kherson, and Zaporizhzhia regions as new subjects of the Russian Federation, often referencing legal status, administrative changes, and ongoing disputes with Ukraine.

2. Ongoing Territorial Dispute with Ukraine: Content is added to highlight Russia's claims over disputed territories, the continuation of Ukrainian contestation, and the legal framing of these regions' status within Russia.

3. Administrative and Municipal Reorganization: Numerous additions detail the restructuring of local governance, including the creation, merging, or abolition of settlements, districts, and municipal entities, especially in newly integrated or reorganized regions.

4. Legal and Constitutional Justification: Additions stress the legal basis for territorial changes, referencing Russian laws, constitutional norms, and the prioritization of Russian legal frameworks over others.

5. Suppression of LGBT Rights and Propaganda: Content is added to reinforce anti-LGBT narratives, including bans on LGBT propaganda, restrictions on same-sex marriage and adoption, and the criminalization of related activities.

6. Recognition of Russian Honors and Academic Achievements: Additions highlight Russian state awards, academic titles, and scientific conferences, often emphasizing achievements within Russian institutions and the importance of Russian scientific and cultural contributions.

7. Emphasis on Russian Identity and International Position: Content is added to assert Russian identity, international alliances, and Russia's global standing, often referencing international comparisons, Russian sports, and the country's role in world affairs.

Notes: This table reproduces verbatim the output of the GPT-4.1 narrative-extraction pipeline with temperature set to 0. The model's original wording, narrative labels, and overall assessment have not been edited and should not be interpreted as author-written exposition. The within-Ruwiki comparison examines changes in Ruwiki between the fork date (May 1, 2023) and January 1, 2025. This analysis uses the same set of 56,835 articles as the cross-sectional analysis, tracking their within-Ruwiki changes. The LLM identified seven deletion narratives and seven insertion narratives from the 500 terms with the highest χ^2 statistics in each direction, using bigrams and trigrams only.

H Deletion-LASSO: Methodology

We estimate LASSO-regularized logistic regression models to identify article characteristics that predict deletion. Specifically, we consider four alternative representations of article content: raw Wikipedia category labels, which are topic tags assigned to each article; terms extracted from category labels; terms extracted from article titles; and the combination of title and category terms. These representations capture different aspects of article content: category labels reflect Wikipedia's taxonomic organization, whereas titles convey an article's primary subject.

For the text-based representations, we construct document–term matrices containing frequency counts of bigrams and trigrams extracted from article titles and category labels.[34] These n-gram features allow us to identify multi-word phrases associated with topics targeted for deletion. Because the feature space is high-dimensional, potentially containing thousands of terms and category labels, we use LASSO regularization to limit overfitting and perform automatic variable selection. The penalty shrinks the coefficients of less informative features, setting many of them exactly to zero while retaining the strongest predictors of deletion.

Predicting article deletions in our setting is challenging because Ruwiki editors deleted only a very small share (about 0.14%) of all Russian-language Wikipedia articles. The appropriate composition of the control group is therefore not obvious: deleted articles could be compared with all surviving articles, recently created articles, or articles with similar characteristics. Using all non-deleted articles would produce a severely imbalanced sample, with more than 99% of observations serving as controls, potentially overwhelming the signal from deleted articles and creating substantial computational demands. Rather than impose an arbitrary control-group definition, we repeatedly draw bootstrap samples of 5,000 control articles with replacement and estimate a separate LASSO model for each draw. This bootstrap procedure is not intended to generate standard errors or confidence intervals for individual coefficients. Instead, it identifies terms that consistently receive nonzero coefficients across different realizations of the control group, allowing us to isolate stable patterns rather than artifacts of a particular control-group selection. We exclude terms that appear in fewer than five percent of the bootstrap iterations, retaining only features whose predictive power persists across alternative sets of control articles. The exhibits in the main text and this appendix apply a stricter reporting threshold, showing only terms with estimated coefficients of at least 0.05 in at least 75% of the 100 iterations.

For each bootstrap iteration $b = 1, \ldots, 100$, we draw a random sample of 5,000 control articles from the set of non-deleted articles and combine them with all deleted articles to form the estimation sample S_b. The dependent variable y_i equals one if article i was deleted and zero otherwise. The independent variables consist either of document–term matrices constructed from article titles and category labels or of binary indicators for membership in individual Wikipedia categories. These variables form a design matrix $\mathbf{X}$ with n observations and p features, where p may exceed n.

Specifically, the resulting LASSO logistic regression estimates coefficients $\boldsymbol{\beta}$ by minimizing the penalized negative log-likelihood:

$$\min_{\boldsymbol{\beta}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i) \right] + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

where $\pi_i = \Pr(y_i = 1 \mid \mathbf{x}_i) = \frac{\exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})}$ represents the predicted deletion probability for article i,

[34] N-grams are sequences of consecutive words. A bigram consists of two consecutive words (e.g., human rights"), and a trigram consists of three consecutive words (e.g., international human rights"). N-grams capture multi-word concepts that individual words may not represent.

and $\lambda \geq 0$ controls the strength of the L1 penalty applied to the coefficient vector.[35]

The estimated coefficients $\hat{\beta}_j$ from the LASSO logistic regression represent log-odds ratios: a one-unit increase in feature j (e.g., the presence of a particular n-gram) changes the predicted log-odds of deletion by $\hat{\beta}_j$, holding all other features constant. Exponentiating yields odds ratios: $\exp(\hat{\beta}_j)$ indicates the multiplicative change in the predicted odds of deletion associated with feature j. However, the L1 penalty biases coefficient estimates toward zero: LASSO shrinks nonzero coefficients relative to unpenalized maximum-likelihood estimates. This shrinkage trades increased bias for reduced variance, improving prediction by limiting overfitting, but generally attenuates the estimated strength of the relationship between individual features and deletion. The LASSO coefficients therefore provide conservative measures of predictor importance, potentially understating genuine relationships while helping to exclude spurious associations.

We retain only features selected with nonzero coefficients in at least 5% of the bootstrap samples, thereby focusing on terms whose predictive power persists across alternative realizations of the control group. The reported exhibits apply a stricter threshold, showing only terms selected with coefficients of at least 0.05 in at least 75% of the iterations. After completing the statistical analysis, we translate the identified Russian terms into English to facilitate interpretation and presentation.

I Deletion-LASSO: Additional Results

This appendix section presents alternative specifications of the deletion-prediction LASSO model of Section VI.A.1, varying the feature space and the inclusion of unigrams.

[35] We estimate the models using the `gamlr` package in R, which implements efficient coordinate-descent algorithms for high-dimensional regularized regression (Taddy, 2017). The penalty parameter λ governs the degree of shrinkage: larger values produce sparser models with fewer nonzero coefficients but may exclude relevant predictors, whereas smaller values retain more features but increase the risk of overfitting. We select λ using five-fold cross-validation. In each bootstrap sample, we partition the observations into five folds, estimate the model on four folds, and evaluate out-of-sample deviance on the remaining fold. We repeat this procedure across a sequence of candidate penalty values and select the value that minimizes average out-of-sample deviance across folds, denoted λ_{min}.

Table A17: LASSO Deletion Prediction Results: Top 15 Predictors by Specification (Part 1)

Edit Notice Terms Included				Edit Notice Terms Filtered			
#	Term	Coef.	Broad topic	#	Term	Coef.	Broad topic
Article Categories							
1	Pornography articles	6.51	Culture	1	Pornography articles	6.51	Culture
2	year in Ukraine (calendar year, general context)	4.49	Other	2	year in Ukraine (calendar year, general context)	4.49	Other
3	articles with editnotice	3.45	Other	3	articles with editnotice	3.45	Other
Category Terms							
1	pornography statute (legal article)	6.73	Other	1	pornography statute (legal article)	6.70	Other
2	LGBT rights movement (emphasis on activism)	4.57	Other	2	LGBT rights movement (emphasis on activism)	4.53	Other
3	invasion by Russia	4.04	Other	3	invasion by Russia	4.11	Other
4	invasion of Ukraine	3.20	Other	4	invasion of Ukraine	3.13	Other
5	caution editing	2.84	War	5	Russo- (prefix for wars involving Russia as a state, e.g., Russo-Turkish War)	3.12	Other
6	Russo- (prefix for wars involving Russia as a state, e.g., Russo-Turkish War)	2.79	Other	6	battle territory	2.49	War
7	war crime	1.65	War	7	war crime	1.81	War
8	battle territory	1.35	War	8	Russian Federation (incorrect grammatical form, used colloquially or in error)	0.37	Other
9	English language (general term)	0.35	Other	9	English language (general term)	0.32	Other
Title Terms							
1	invasion by Russia	5.26	Other	1	invasion by Russia	5.26	Other
2	pornography statute (legal article)	4.65	Other	2	pornography statute (legal article)	4.65	Other
Category Terms (with unigrams)							
1	pornography (general term)	6.26	Other	1	pornography (general term)	5.79	Other
2	sexual	3.96	Culture	2	sexual	3.96	Culture
3	LGBT people (demographic group)	3.19	Other	3	LGBT people (demographic group)	3.20	Other
4	activist (general)	2.01	Other	4	activist (general)	2.01	Other
5	invasion (general term)	1.92	Other	5	invasion (general term)	1.90	Other
6	Russian (relating to the Russian Federation as a state or its citizens)	1.65	Other	6	Russian (relating to the Russian Federation as a state or its citizens)	1.75	Other
7	caution	1.38	War	7	Donetsk	1.45	War
8	Donetsk	1.37	War	8	public	1.36	Politics
9	public	1.35	Politics	9	person (human being)	1.23	Other
10	person (human being)	1.21	Other	10	Ukraine (country name, nominative case)	1.15	Other
11	Ukraine (country name, nominative case)	1.15	Other	11	woman (adult female person)	0.83	Other
12	mass-scale (large-scale, widespread)	0.83	Other	12	mass-scale (large-scale, widespread)	0.81	Other
13	woman (adult female person)	0.83	Other	13	article	0.75	Other
14	war crime	0.73	War	14	Russia (geographical/historical region)	0.58	Other
15	Russia (geographical/historical region)	0.57	Other	15	political	0.58	Politics
Title Terms (with unigrams)							
1	pornography (general term)	4.12	Other	1	pornography (general term)	4.12	Other
2	skirmish (small-scale fight)	3.57	Other	2	skirmish (small-scale fight)	3.57	Other
3	Ukraine (country name, nominative case)	2.74	Other	3	Ukraine (country name, nominative case)	2.74	Other
4	Russian (adjective)	2.60	Politics	4	Russian (adjective)	2.60	Politics
5	Russia (geographical/historical region)	1.58	Other	5	Russia (geographical/historical region)	1.58	Other

Notes: This table presents the top 15 deletion predictors from bootstrapped LASSO logistic regression for each specification. For each term, the reported coefficient is the mean of its estimated LASSO coefficient across the bootstrap iterations in which the term is selected (i.e., conditional on a nonzero coefficient); iterations in which the term is not selected do not enter the average. Coefficients represent log-odds ratios systematically shrunk toward zero by L1 regularization. Terms selected with per-iteration coefficients of at least 0.05 in fewer than 75% of bootstrap iterations are excluded. Default mode includes edit notice terms; No-Editnotice mode filters these terms during n-gram generation. Categories: Culture (pornography, LGBTQIA+ topics), War (Ukraine invasion, conflict), Politics (Russian political figures, activists), Other (miscellaneous).

Table A18: LASSO Deletion Prediction Results: Title+Category Terms with Unigrams

Edit Notice Terms Included				Edit Notice Terms Filtered			
#	Term	Coef.	Broad topic	#	Term	Coef.	Broad topic
Title+Category Terms (with unigrams)							
1	pornography (general term)	5.83	Other	1	pornography (general term)	5.44	Other
2	sexual	3.26	Culture	2	sexual	3.26	Culture
3	LGBT people (demographic group)	2.41	Other	3	LGBT people (demographic group)	2.42	Other
4	activist (general)	2.17	Other	4	activist (general)	2.18	Other
5	invasion (general term)	1.29	Other	5	invasion (general term)	1.28	Other
6	caution	1.17	War	6	Russian (relating to the Russian Federation as a state or its citizens)	1.13	Other
7	Russian (relating to the Russian Federation as a state or its citizens)	1.06	Other	7	Donetsk	1.10	War
8	Donetsk	1.06	War	8	public	1.04	Politics
9	public	1.04	Politics	9	person (human being)	1.00	Other
10	Ukraine (country name, nominative case)	0.99	Other	10	Ukraine (country name, nominative case)	0.99	Other
11	person (human being)	0.98	Other	11	Russo- (prefix for wars involving Russia as a state, e.g., Russo-Turkish War)	0.82	Other
12	Russo- (prefix for wars involving Russia as a state, e.g., Russo-Turkish War)	0.80	Other	12	article	0.76	Other
13	war crime	0.69	War	13	war crime	0.68	War
14	mass-scale (large-scale, widespread)	0.69	Other	14	mass-scale (large-scale, widespread)	0.68	Other
15	woman (adult female person)	0.63	Other	15	woman (adult female person)	0.64	Other

Notes: This table presents the top 15 deletion predictors from bootstrapped LASSO logistic regression for each specification. For each term, the reported coefficient is the mean of its estimated LASSO coefficient across the bootstrap iterations in which the term is selected (i.e., conditional on a nonzero coefficient); iterations in which the term is not selected do not enter the average. Coefficients represent log-odds ratios systematically shrunk toward zero by L1 regularization. Terms selected with per-iteration coefficients of at least 0.05 in fewer than 75% of bootstrap iterations are excluded. Default mode includes edit notice terms; No-Editnotice mode filters these terms during n-gram generation. Categories: Culture (pornography, LGBTQIA+ topics), War (Ukraine invasion, conflict), Politics (Russian political figures, activists), Other (miscellaneous).

J Examples of Edits

This appendix provides translated difference visualizations of representative heavily edited articles, illustrating the systematic patterns presented in Section VI.B. In each figure, text segments highlighted in green were added to Ruwiki (relative to the Wikipedia version on the same day), while text segments highlighted in red were deleted from Wikipedia (i.e., removed in the Ruwiki version). The differences are based on comparing the same articles across platforms as of January 1, 2025 (twenty months after the fork). The examples are deliberately drawn from articles whose subjects fall squarely within the three pillars of Section V.

Figure A7: Example of heavy edits: *LGBT rights in Russia* article (cultural pillar).

Observance of the rights and freedoms of Russian citizens belonging to the LGBT (homosexual, bisexual and transgender people) is considered in the West a pressing problem in modern Russia. The invisibility and infringement of representatives The propaganda of same-sex relations in Russia is prohibited by law. Most Russian lawyers and human rights activists believe that LGBT representatives in Russia are perceived as a normal phenomenon. This is associated with manifestations of aggression, intolerance and political repression against LGBT in Russia. want special rights for themselves.

Figure A8: Example of heavy edits: *The Katyn massacre* article (foreign-policy and historical pillar).

The Katyn massacre (sometimes also shortened to Katyn - the Katyn crime) was a war crime, mass murder, mass execution of Polish citizens, mainly captured officers of the Polish army, carried out in the spring of 1940 , prisoners of war and civilians, committed in the Katyn forest in the Smolensk region. To this day, the question of who was guilty of this crime remains the subject of scholarly and public debate. According to the first point of view, based on the investigation of the German occupation authorities in 1943. The executions were carried out and according to Soviet documents declassified in 1992 by decision of a special "troika" on the order of Boris Yeltsin, responsibility for the execution is assigned to the NKVD of the USSR in accordance with the resolution of the Politburo of the Central Committee of the All-Union Communist Party (Bolsheviks) of March 5, 1940; executed . According to the second point of view, which was held by the Soviet side of the prosecution at the International Military Tribunal in Nuremberg, the shooting was carried out by Nazi Germany.

Figure A9: Example of heavy edits: *Vladimir Putin* article (domestic-politics pillar).



Figure A10: Example of heavy edits: *Human rights in Russia* article (domestic-politics pillar).



Figure A11: Example of heavy edits: *Aleksei Navalny* article (domestic-politics pillar).



Figure A12: Example of heavy edits: *1984* article (domestic-politics and foreign policy pillars: application to history and literature).



K Editing-LASSO Details

In this appendix section, we provide details of the methodology underlying the results of the phrase-level LASSO analysis reported in Section VI.B.2 of the main text. The empirical specification is presented in Subsection K.1 and the conceptual framework that helps interpret the results in Subsection E.

K.1 Empirical Specification

This appendix section provides the technical details underlying the phrase-level LASSO analysis reported in Section VI.B.2. The goal of the exercise is to identify the bigrams and trigrams that systematically distinguish text inserted into Ruwiki from text removed relative to Wikipedia, conditional on comparing changes within the same article.

Empirical specification. For each text segment i in article p, let $y_i = 1$ if the segment is an insertion and $y_i = 0$ if it is a deletion. Let X_i be a vector of binary indicators for the presence of each lemmatized bigram or trigram. We estimate the LASSO logistic regression

$$\Pr(y_i = 1 \mid X_i, \alpha_p) = \frac{1}{1 + \exp[-(\alpha_{p[i]} + X_i'\beta)]},$$

where α_p are article-specific intercepts (article effects) and β is the coefficient vector of interest. The estimates solve the penalized maximum-likelihood problem

$$(\hat{\alpha}, \hat{\beta}) = \arg\max_{\alpha,\beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} [y_i \log p_i + (1 - y_i) \log(1 - p_i)] - \lambda \left(\sum_{p=1}^{P} |\alpha_p| + \sum_{j=1}^{J} |\beta_j| \right) \right\}.$$

Both the article effects α_p and the n-gram coefficients β are penalized under a common L1 penalty, with λ selected by 10-fold cross-validation on out-of-sample deviance. The regularized article effects flexibly absorb article-level differences in editing intensity and topic: $\hat{\alpha}_p$ is nonzero only for articles with a strongly imbalanced mix of insertions and deletions, so that $\hat{\beta}_j$ captures the systematic association between an n-gram and the *direction* of editorial change *within* an article. Penalizing the article effects also rules out perfect separation for articles containing only insertions or only deletions and avoids the incidental-parameters problem of unpenalized fixed-effects logit. The estimated coefficients are shrunken log-odds used for descriptive term selection, not structural parameter estimates.

K.2 Applying the Conceptual Framework to the Editing-LASSO Analysis

This subsection explains how we interpret the baseline comparison between Ruwiki and Wikipedia and the placebo comparison within-Wikipedia in the editing-LASSO analysis using the conceptual framework developed in Online Appendix Section E.

In the LASSO analyses of deletions and insertions within heavily edited articles, we compare the cross-platform baseline with the within-Wikipedia placebo. Combining the two comparisons gives rise to four conceptually distinct groups of terms.

Taxonomy of terms selected by the editing LASSO.

(i) **Terms selected in both comparisons with opposite signs: staleness or censorship by omission.** Terms selected in both the R_T vs. W_T and the W_T vs. W_F comparisons are naturally interpreted as

reflecting staleness. These are terms whose prevalence changed within Wikipedia over time and whose cross-platform gap can therefore be explained by Ruwiki's failure to incorporate Wikipedia's updates. In the frozen-clone benchmark, in which Wikipedia moves with the world while Ruwiki remains in its fork-date state, the two comparisons would select the same terms with opposite signs: terms classified as deletions in one comparison would be classified as insertions in the other, and vice versa. Some of these terms may nevertheless reflect censorship by omission when Ruwiki selectively fails to incorporate new information that is sensitive to the regime. Empirically, however, we cannot distinguish such selective non-updating from passive staleness.

(ii) **Terms selected only in the cross-platform baseline comparison: active editing.** Terms selected only in the cross-platform comparison are the most direct markers of active Ruwiki divergence. Because R_T and W_T are observed at the same date, these terms cannot be explained by changes in the world between F and T. Moreover, under the frozen-clone benchmark, a purely stale term should also appear in the within-Wikipedia comparison, since the cross-platform gap would mirror the temporal change within Wikipedia. Thus, terms that distinguish R_T from W_T but do not distinguish W_T from W_F are naturally interpreted as the linguistic footprint of active Ruwiki editing. Terms overrepresented in W_T relative to R_T identify content that Ruwiki removed or suppressed; terms overrepresented in R_T relative to W_T identify content that Ruwiki inserted or amplified.[36]

(iii) **Terms selected only in the within-Wikipedia placebo: change in the world and routine editorial drift.** These terms distinguish Wikipedia in January 2025 from Wikipedia at the fork, but they do not strongly distinguish Ruwiki from Wikipedia in January 2025. They therefore capture changes that occurred within Wikipedia over time rather than changes distinctive to Ruwiki. Such terms may reflect new events, new vocabulary, or routine editorial updating. The fact that they do not appear in the cross-platform comparison suggests that Ruwiki either incorporated the same updates or that any remaining cross-platform difference is small relative to the stronger signals of active Ruwiki editing. These terms are therefore useful as a benchmark: they show what ordinary evolution of Wikipedia content looks like over the same period.

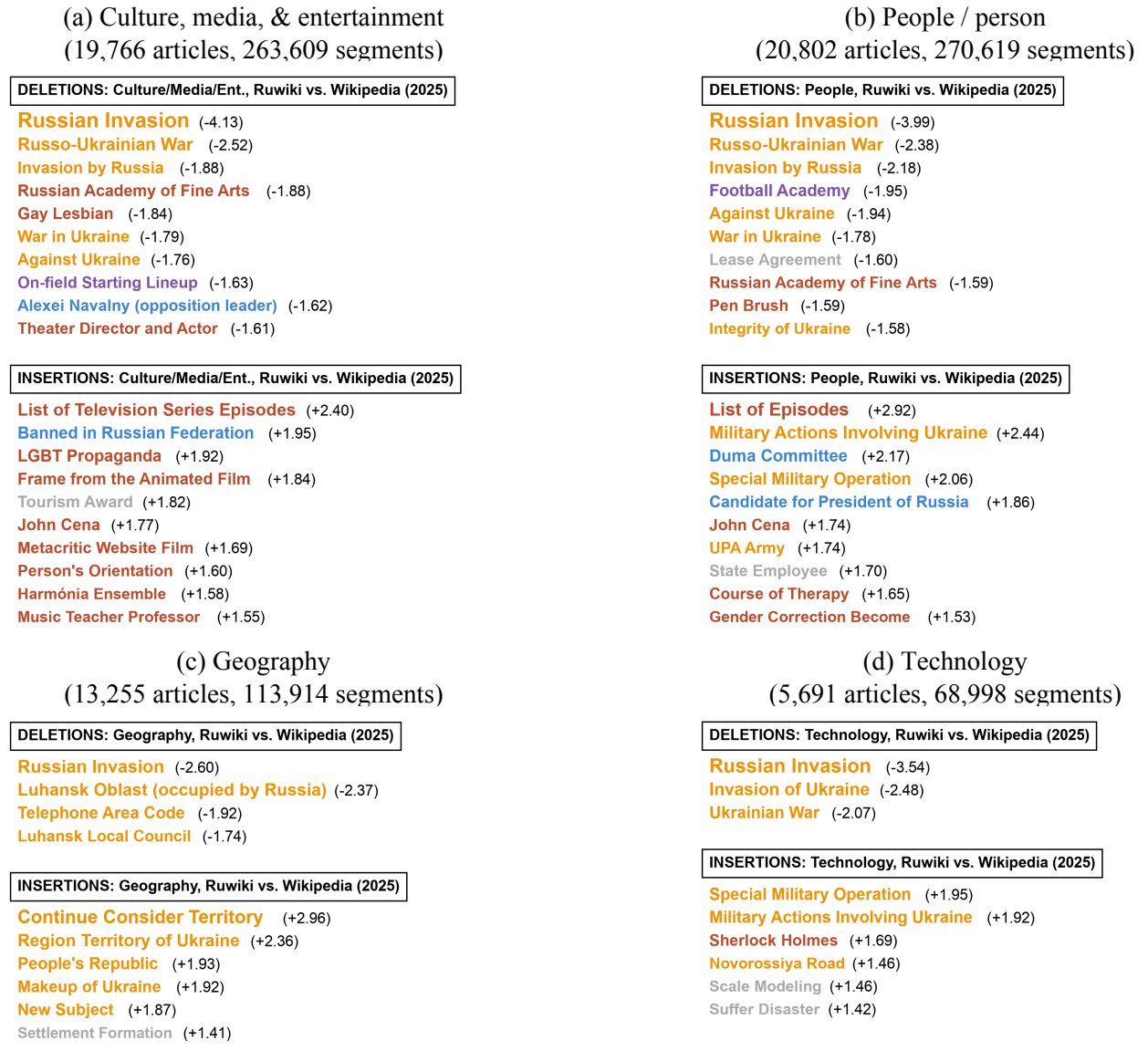
(iv) **Terms selected in both comparisons with the same sign: differential updating.** A term may become more prevalent in Wikipedia in response to a new event and be used even more extensively by Ruwiki at T. In this case, both platforms update in the same direction but with different intensity or emphasis. In practice, this category is empirically negligible.

In Appendix Section L.2 (below), we assess the discriminating power of the selected vocabulary using held-out classification. The vocabulary distinguishes Ruwiki from contemporaneous Wikipedia substantially better than it distinguishes contemporaneous Wikipedia from its fork-date version.

[36] This logic should be interpreted in light of the competitive nature of LASSO selection. A term's absence from one list does not mechanically imply that the corresponding difference is exactly zero. Rather, the placebo comparison provides a demanding benchmark: if a term's cross-platform gap were driven primarily by ordinary temporal change in Wikipedia, we would expect it to have strong predictive content in the W_F vs. W_T comparison as well.

L Editing-LASSO: Additional Results

Figure A13: Baseline editing LASSO, additional LLM topic categories:
Terms selected only in the baseline Ruwiki-Wikipedia comparison in January 2025

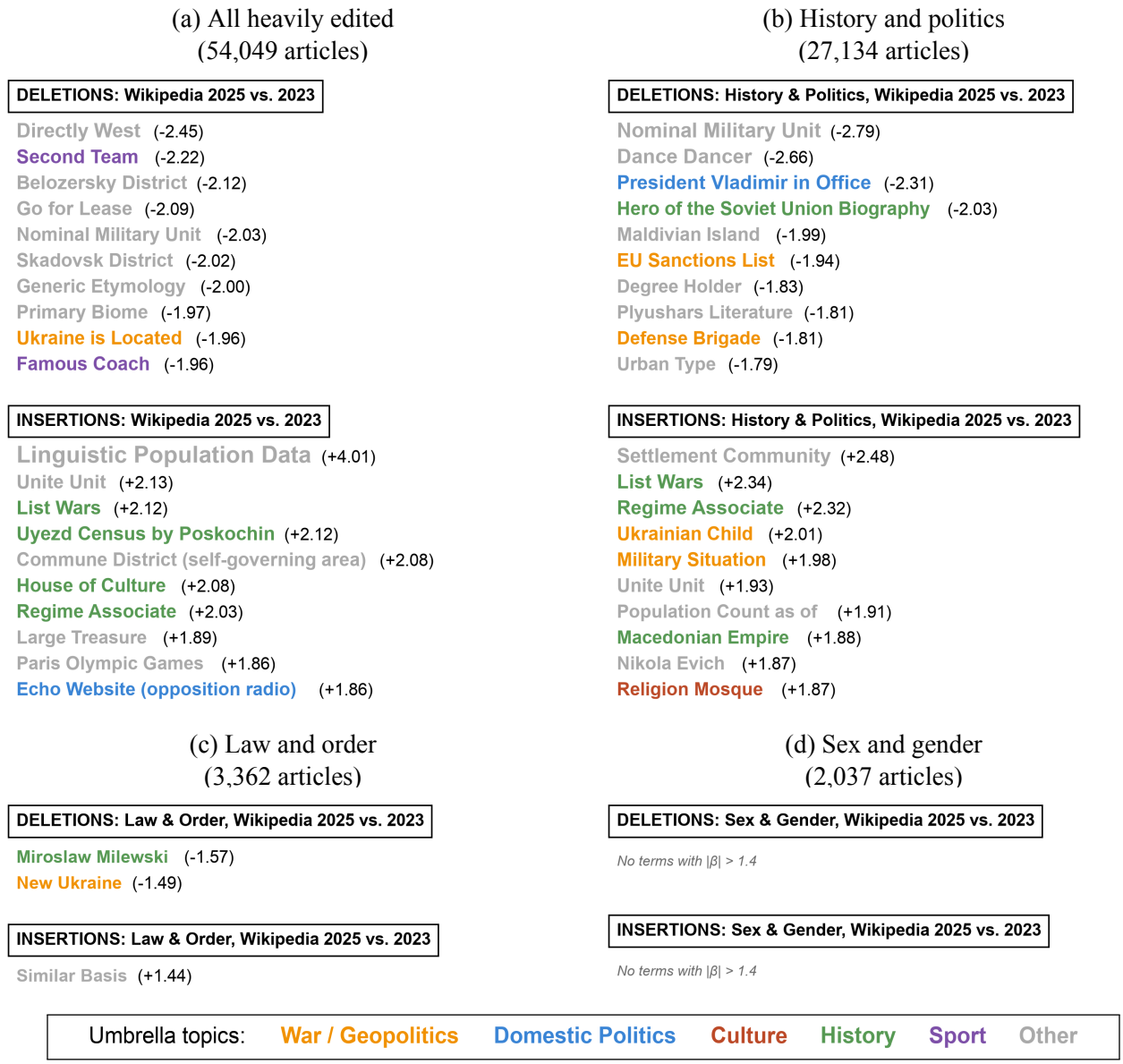


Notes: Terms selected by the LASSO in the January 2025 Ruwiki–Wikipedia comparison but not in the within–Wikipedia placebo. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions, provided that $|\hat{\beta}| > 1.4$. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. No terms shown are classified as HISTORY; SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples: culture, media, and entertainment, 263,609 segments in 19,766 articles; people/person, 270,619 in 20,802; geography, 113,914 in 13,255; and technology, 68,998 in 5,691.

L.1 Within-Wikipedia Temporal Placebo: Editing LASSO

This section presents the within-Wikipedia temporal placebo for the editing LASSO, paralleling the cross-sectional analysis in Section VI.B. The temporal placebo compares Russian-language Wikipedia in May 2023 against Russian-language Wikipedia in January 2025. Figures A14 and A15 present terms chosen by LASSO in within-Wikipedia comparision and not in Ruwiki-Wikipedia baseline. These are pure Wikipedia placebo results. Figures A16 and A17 present terms chosen by LASSO in both within-Wikipedia comparison and in Ruwiki-Wikipedia baseline, where they appear with the opposite sign. These terms represent Ruwiki staleness or censorship by omission. There is only one term chosen in both comparisons with the same sign, this is "koatuu code" (old Ukrainian administrative code), which predicts deletions in Ruwiki and Wikipedia. As explained in Appendix Section E, selection in the same direction reflects updating of different intensity in response to the same real-world event. Here, the relevant event was the replacement of outdated Ukraine's KOATUU administrative classification code with the correct KATOTTH code.

Figure A14: LASSO predictions of deletions and insertions within-Wikipedia in heavily edited articles: Terms selected only in the within-Wikipedia placebo



Notes: Terms selected by the LASSO in the within–Wikipedia placebo between 2023 and 2025, but not in the January 2025 Ruwiki–Wikipedia baseline comparison. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions, provided that $|\hat{\beta}| > 1.4$. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples.

Figure A15: Placebo within-Wikipedia editing LASSO, additional topics:
Terms selected only in the within-Wikipedia placebo

(a) Culture, media & entertain-t (19,099 articles)

DELETIONS: Culture/Media/Ent., Wikipedia 2025 vs. 2023

- Playlist (-2.02)
- Second Team (-1.92)
- Lev Perfil (-1.77)
- Sant Maria (-1.60)
- Game Review (-1.56)
- Reference Dictionary (-1.55)
- Dubbed Roles (-1.44)

INSERTIONS: Culture/Media/Ent., Wikipedia 2025 vs. 2023

- Starting Field (+1.76)
- Reviewers Opinion (+1.62)
- Yandex Music (streaming service) (+1.51)
- Digital Release (+1.45)
- Multiple Prizewinner (+1.45)
- Roland Garros (+1.43)
- Original Painting (+1.42)

(b) People / person (20,802 articles)

DELETIONS: People, Wikipedia 2025 vs. 2023

- Hero of the Soviet Union Biography (-2.74)
- President Vladimir in Office (-2.58)
- Scientific Works (-2.34)
- Boxing Prospect (-2.06)
- Sant Maria (-1.86)
- Literature Primary Source (-1.82)
- Official Code (-1.79)
- Mikhail Chigorin (chess player) (-1.76)
- Reference Dictionary (-1.72)
- Assistant Director (-1.71)

INSERTIONS: People, Wikipedia 2025 vs. 2023

- List Wars (+2.55)
- Paris Olympic Games (+2.54)
- Regime Associate (+2.32)
- European Union (+2.22)
- General Seniority Correction (+1.99)
- Medalist Zone (+1.97)
- Andreyevich (+1.94)
- Footballer Become (+1.92)
- Nikola Evich (+1.81)
- Aleksandrovich (+1.77)

(c) Geography (13,255 articles)

DELETIONS: Geography, Wikipedia 2025 vs. 2023

- Directly West (-2.76)
- Local Council (-2.53)
- Belozersky District (-2.52)
- Skadovsk District (-2.47)
- Slavyansky District (-2.21)
- Urban Type (-2.18)
- Main Century (-2.14)
- Tatar Institute (-2.09)
- East Direct (-2.04)
- Henichesk District (-2.03)

INSERTIONS: Geography, Wikipedia 2025 vs. 2023

- Settlement Community (+2.50)
- Commune District (self-governing area) (+2.49)
- Unite Unit (+2.29)
- Urban Settlement (+2.18)
- Estate Development (+2.18)
- Uyezd Census by Poskochin (+2.11)
- Population Residing as Follows (+1.96)
- System State (+1.87)
- House of Culture (+1.86)
- Belen Ex. (+1.85)

(d) Technology (5,691 articles)

DELETIONS: Technology, Wikipedia 2025 vs. 2023

- Game Review (-1.58)
- Game Feature (-1.41)

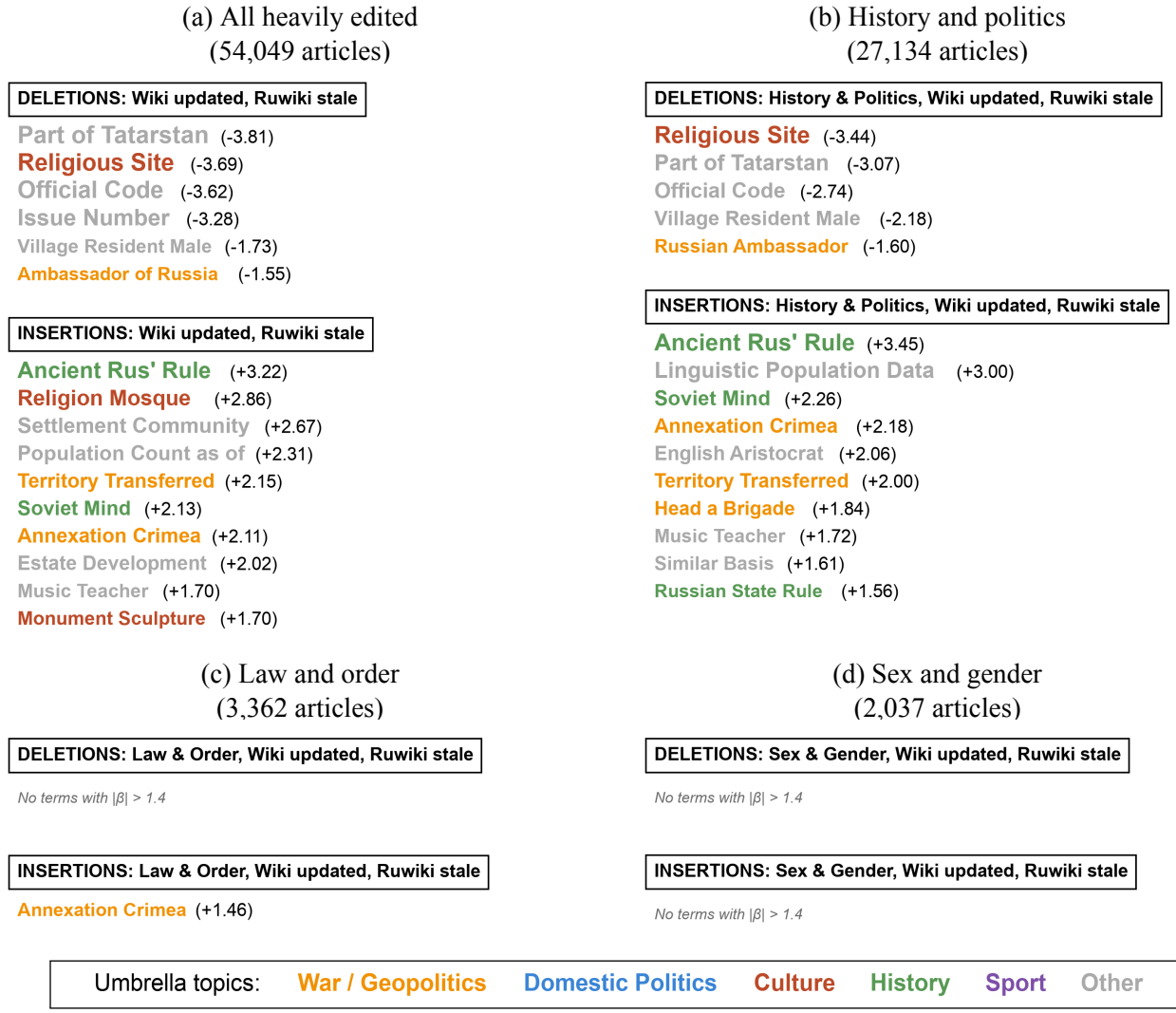
INSERTIONS: Technology, Wikipedia 2025 vs. 2023

No terms with $|\beta| > 1.4$

Umbrella topics: War / Geopolitics | Domestic Politics | Culture | History | Sport | Other

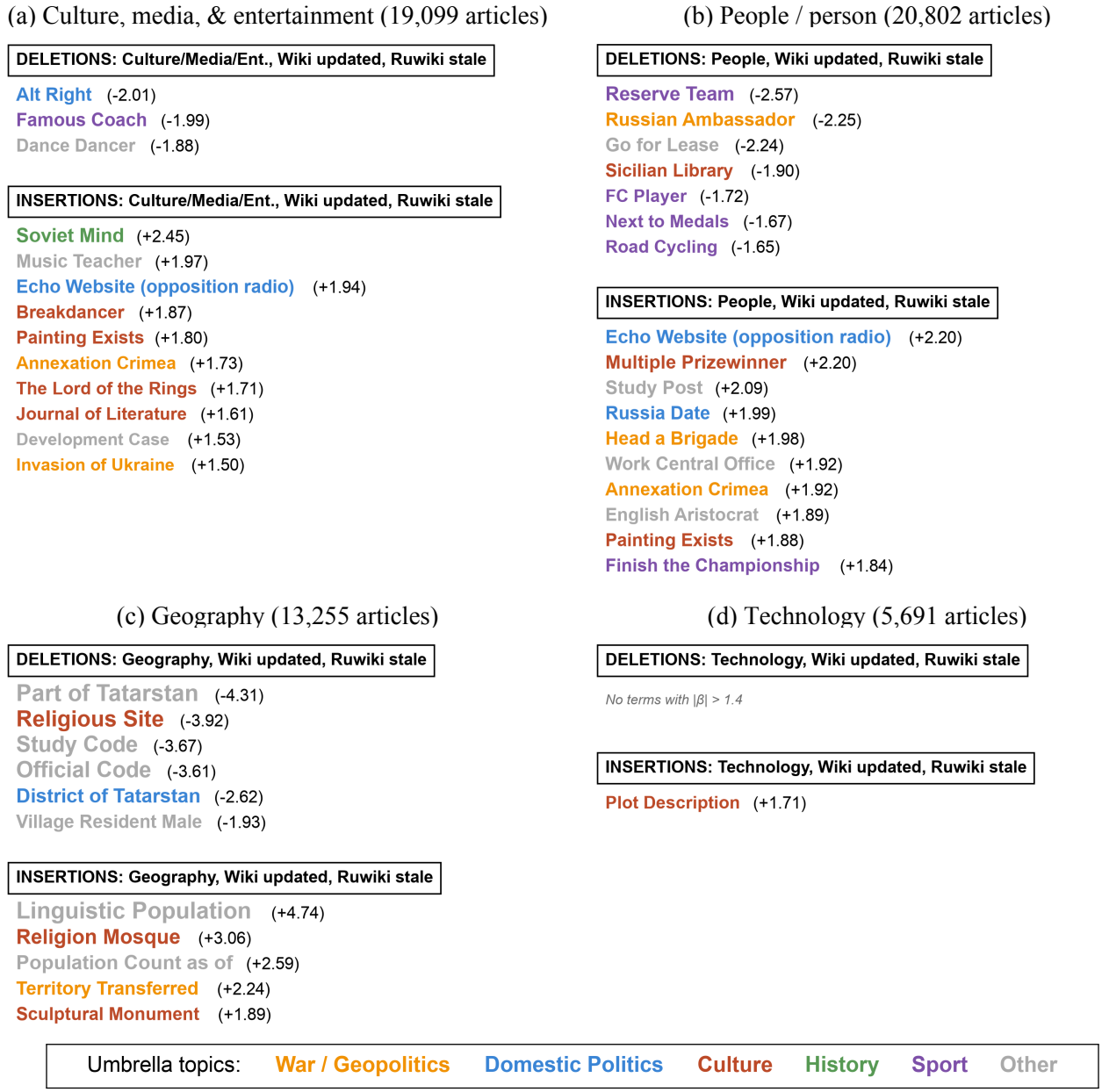
Notes: Terms selected by the LASSO in the within–Wikipedia placebo between 2023 and 2025, but not in the January 2025 Ruwiki–Wikipedia baseline comparison. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions, provided that $|\hat{\beta}| > 1.4$. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples.

Figure A16: LASSO: Wikipedia updating and Ruwiki staleness and passive censorship
Terms selected in both baseline and placebo with the opposite sign



Notes: Terms selected by the LASSO in both the January 2025 Ruwiki–Wikipedia baseline comparison and the within–Wikipedia placebo between 2023 and 2025. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions, provided that $|\hat{\beta}| > 1.4$. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples.

Figure A17: LASSO: Wikipedia updating and Ruwiki staleness and passive censorship
Additional topics: Terms selected in both baseline and placebo with the opposite sign



Notes: Terms selected by the LASSO in both the January 2025 Ruwiki–Wikipedia baseline comparison and the within–Wikipedia placebo between 2023 and 2025. The figure reports the ten terms with the largest absolute coefficients predicting additions or deletions, provided that $|\hat{\beta}| > 1.4$. Coefficients, ($\hat{\beta}_{\text{term}}$), come from regularized logistic regressions in which positive values predict additions and negative values predict deletions. Terms are ranked by ($|\hat{\beta}|$), with font size increasing indicatively, rather than proportionally, with coefficient magnitude; exact coefficients are shown next to each term. Models are estimated separately by topic using lemmatized bigrams and trigrams, article fixed effects, and 10-fold cross-validation. Colors denote manually coded semantic umbrella categories. SPORT is an additional category beyond the five used for deleted terms. The unit of observation is a text segment. Because the document–term matrix is sparsity-filtered before estimation, estimation samples are slightly smaller than the full samples.

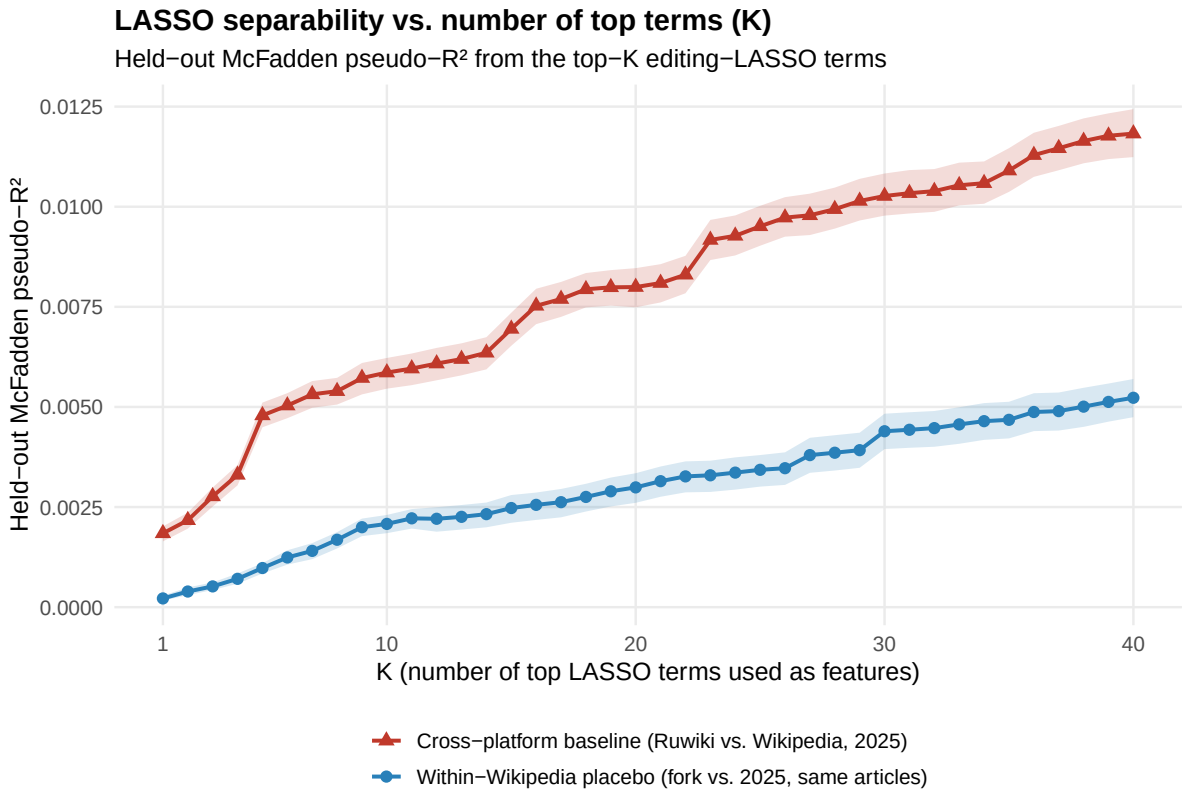
L.2 Competitive LASSO selection and separability

Selection into each LASSO list is competitive: a penalized model retains a term only if it has predictive value relative to correlated alternatives within the same comparison. Moreover, the two comparisons differ in difficulty. The cross-platform comparison contains strong discriminators generated by active editing of Ruwiki, whereas the within-Wikipedia comparison contrasts two highly similar snapshots of the same encyclopedia. The within-Wikipedia placebo may therefore select weak or idiosyncratic terms simply because stronger discriminators are absent.

For this reason, we compare not only the composition of the selected term lists but also the overall separability of the two classification exercises. For each comparison, we take the top K terms ranked by the corresponding editing LASSO and use them to fit a logistic classifier of snapshot membership: R_T versus W_T in the baseline comparison and W_F versus W_T in the placebo comparison. We then report the held-out McFadden pseudo-R^2. Holding K and the train–test procedure fixed makes the fit measures directly comparable. A higher held-out pseudo-R^2 in the cross-platform comparison indicates that Ruwiki and Wikipedia are more readily distinguishable in January 2025 than Wikipedia is from its own fork-date version. This provides a summary measure of the strength of active Ruwiki divergence relative to ordinary temporal change within Wikipedia.

Appendix Figure A18 reports the results for $K = 1, \ldots, 40$. We use the same 70/30 train–test split in both comparisons. The baseline curve lies above the placebo curve at every K. The relative difference is largest among the most predictive terms: at $K = 1$, the baseline pseudo-R^2 is approximately 5.9 times the placebo value; the ratio narrows to roughly 2.3 by $K = 40$. Both held-out pseudo-R^2 curves increase monotonically over the displayed range. This monotonicity is not mechanical, since adding predictors need not improve held-out fit, and indicates that lower-ranked terms continue to contribute out-of-sample discriminatory power. We also report 95% confidence intervals from an article-clustered bootstrap. The baseline and placebo intervals do not overlap at any K in the displayed range. Thus, across the top-ranked terms, the selected vocabulary separates Ruwiki from contemporaneous Wikipedia roughly two to three times more sharply than it separates Wikipedia from its own fork-date version. This pattern is consistent with the cross-platform differences reflecting active Ruwiki editing rather than ordinary updating of Wikipedia over the same period. The absolute pseudo-R^2 values are small because most edited segments contain none of the selected terms; the informative quantity is therefore the relative fit across the two otherwise identical classification exercises.

Figure A18: Separability of LASSO Ruwiki to Wikipedia vs. within-Wikipedia comparisons



Note: Held-out McFadden pseudo-R^2 of a logistic classifier of snapshot membership fit on the top-K terms ranked by the editing LASSO, as a function of K ($K = 1, \ldots, 40$). The unit of observation is the individual edited segment (a single added or deleted text passage within a heavily edited article); segments are nested within articles. The cross-platform baseline (red triangles) predicts R_T vs. W_T (Ruwiki vs. Wikipedia, January 2025); the within-Wikipedia placebo (blue circles) predicts W_F vs. W_T (Wikipedia at the fork, May 2023, vs. January 2025). Features are binary term incidences; K and the 70/30 train–test split are held fixed across the two comparisons. Shaded bands are 95% confidence intervals from an article-clustered bootstrap (300 resamples of whole articles, resampled at the article level because edited segments are not independent within an article). Higher values mean the top-K terms separate the two snapshots more sharply.

M Editing Robustness: Cramér's V analyses

Table A19: Terms most characteristic of the Ruwiki–Wikipedia baseline comparison only (signed Cramér's V)

All heavily edited	
Deletions	*Insertions*
postal code, telephone index, postal code telephone, telephone area code, telephone prefix code, Russo-Ukrainian War, Ukrainian war, integrity of Ukraine, KOATUU local, KOATUU local council, Luhansk Oblast, Luhansk local council, Luhansk Council, council of Luhansk Oblast, Alexei Navalny, KOATUU local code, telephone exchange code, Novaya Gazeta, telephone code KOATUU, sexual orientation	new subject, part Russian Federation, subject composition, composition of Russian subjects, new subject composition, Russian Federation defines, federation defines, Russian Federation authority, federation authority, Ukrainian authority disputed, Ukraine disputes, Russian composition, Ukraine continues to dispute, define territory, the government of Ukraine, federation authority Ukraine, including village, continue consider territory, consider territory, people's republic
History & politics	
Deletions	*Insertions*
telephone index, postal code telephone, postal code, telephone area code, telephone prefix code, local council, integrity of Ukraine, Russo-Ukrainian War, Ukrainian war, Alexei Navalny, Luhansk Oblast, KOATUU local, KOATUU local council, Luhansk local council, Luhansk Council, council of Luhansk Oblast, KOATUU local code, Novaya Gazeta, sexual orientation, gay lesbian	new subject, part Russian Federation, composition of Russian subjects, subject composition, new subject composition, Russian Federation defines, federation defines, Russian Federation authority, federation authority, Ukrainian authority disputed, Ukraine disputes, define territory, Russian composition, Ukraine continues to dispute, continue consider territory, federation defines territory, consider territory, federation authority Ukraine, including village, the government of Ukraine
Culture, media, & entertainment	
Deletions	*Insertions*
Alexei Navalny, Novaya Gazeta, gay lesbian, Russian state authority, start of invasion, war in Ukraine, Echo Moscow, Vladimir Putin, Radio Liberty, Russo-Ukrainian War, Ukrainian war, background of the invasion, to have sexual intercourse, minute of the meeting, background of the Russian invasion, Echo website, freedom of speech, Russia against, territory Ukraine, protest against	military actions involving Ukraine, actions Ukraine, list of television series episodes, military operations, list of episodes, conference proceedings, material presented scientifically, television series installment, LGBT advocacy, scientific compendium, scientific material, Ministry of Justice of the Russian Federation, materials of the scientific conference, propaganda of the LGBT community, banned in Russian Federation, scientific proceedings, scientific symposium, university named after, foreign agent, become a scientific conference
Law & order	
Deletions	*Insertions*
sexual orientation, Alexei Navalny, Novaya Gazeta, Russian state authority, gay lesbian, sexual trait, same-gender couple, sign of sexual orientation, lesbian gay, gender identity, gay bisexual, sexual orientation gender, lesbian gay bisexual, Vladimir Putin, integrity of Ukraine, ground of hatred, point of view, Echo Moscow, judgment of the court, human rights	military actions involving Ukraine, actions Ukraine, banned in Russian Federation, military operations, many countries world, code exists, Family Code exists, have a provision, the code contains provision, priority of the Russian Constitution, Russian priority, marriage adoption child, adoption of children by LGBT, child of LGBT person, Russian constitutional norm, constitutional norm, Russian Constitution, prohibited by Russian Federation recognized, a number of Central countries, LGBT child
Sex & gender	
Deletions	*Insertions*
gay lesbian, lesbian gay, gender identity, gay bisexual, sexual orientation gender, sexual trait, sign of sexual orientation, current era, lesbian gay bisexual, ground of hatred, to come into force, anal sex, gay rights, recognition of same-sex, vote against, invasion by Russia, Russian invasion into Ukraine, sex with men, sexual encounter, Russia–Ukraine	banned in Russian Federation, LGBT advocacy, many countries world, code exists, Family Code exists, have a provision, the code contains provision, priority of the Russian Constitution, Russian priority, Russian constitutional norm, adoption of children by LGBT, Russian Constitution, marriage adoption child, constitutional norm, child of LGBT person, prohibited by Russian Federation recognized, RF recognition, a number of Central countries, Central Asian countries, LGBT child

Note: Within each topic block, the columns list the bigrams/trigrams most characteristic of deletions and of insertions for this comparison, ranked by the signed Cramér's V (equivalently the scaled ϕ coefficient from the 2×2 table of n-gram presence by edit type).

Table A20: Terms most characteristic of the within-Wikipedia placebo comparison only (signed Cramér's V)

All heavily edited	
Deletions	*Insertions*
court abolished, rural populated place, federal law, republic de facto village, de facto republic, law December Federal Law, village is actually located, under United sanctions, sanctions United, sanctions imposed by the United States government, December Federal Law, sub-basin river sub-basin, river sub-basin, river sub-basin absent, famous bearer, as conductor perform, NKR district according to, March under sanctions, President Vladimir in office, half Tatar	part of, Order of Merit, Russian state, system scheme, score the first goal, APG system scheme, Presidium Supreme Soviet, first goal, plant diagram, Nagorno declaration, Nagorno-Karabakh declaration, deputy adopted declaration, adopt Nagorno declaration, Nagorno session, Nagorno regional session, Karabakh Autonomous Region, regional district council, joint Nagorno session, supreme presidium, abolition Nagorny
Culture, media, & entertainment	
Deletions	*Insertions*
as conductor perform, orchestra as conductor, song as orchestra, Eurovision competition, return performer, Eurovision song entry, orchestra quality, song orchestra, Eurovision song competition, Eurovision Song Contest song, Eurovision take place, conductor song orchestra, conductor performed, final point total, announce voting, Noela Kilahan, voting results announced, as conductor, Michel Stoker, final score result	cultural heritage, hit position, hit parade position, science-based fiction, Great War, hit parade, jubilee day, chart parade, hit parade chart, technical staff, Arkady Strugatsky, prize named, good novella, duel against, good prize, prose nomination, literary monograph, anniversary birthday, music producer, film comprise
History & politics	
Deletions	*Insertions*
rural populated place, under United sanctions, sanctions United, sanctions imposed by the United States government, President Vladimir in office, as conductor perform, March under sanctions, orchestra as conductor, Eurovision song entry, song as orchestra, March be located, return performer, orchestra quality, song orchestra, as conductor, Eurovision take place, Eurovision competition, United States America, nominal military unit, Eurovision Song Contest song	Russian state, part of, Order of Merit, economic basis, honorary title, jubilee day, Presidium Supreme Soviet, monument history, supreme presidium, order of voivode, inhabited locality, signal battalion, medal award, friendship of peoples, architect's project, Kingdom of Prussia, peace treaty, cultural heritage, be given a name, many years of work
Law & order	
Deletions	*Insertions*
federal law, December law, June law, February law, law July, people's court, city court, People's Deputy, of the White House, militia staff member, military unit reduced, nominal military unit, part reduced, southeast Ukraine, episode of the season, April law, military unit, under United sanctions, sanctions United, sanctions imposed by the United States government	important roles, St. Stanislaus, burial site, death penalty, Commonwealth of Independents, Commonwealth of Independent States, Leningrad Oblast, law faculty, childhood and youth, February delegate, January February delegate, State Councillor, independent state, February delegate congress, Order of Saint Stanislaus, people's assembly, Russian governorate, governorate of the Russian Empire, St. Vladimir, army service
Sex & gender	
Deletions	*Insertions*
good scene, AVN good scene, AVN Award for Best New Starlet, group scene, women's football club, football club, performance result, Best Group Scene, Energiya Voronezh, AVN good group, women's football team, swimsuit result, goalkeeper coach, participant Russian championship, ice hockey puck, Russian championship history, primary lineup, fictional character, AVN nomination, our time	contribution to development, score series, score in penalty shootout, goal scored in series, ball count, record of goals scored, technical accounting, penalty accounting, penalty technical accounting, penalty shootout accounting, put the ball in the net, penalty shootout, Olympic Games, become sultan, women's association, resign from office, men women, become laureate, technical result accounting, women's movement

Note: Within each topic block, the columns list the bigrams/trigrams most characteristic of deletions and of insertions for this comparison, ranked by the signed Cramér's V (equivalently the scaled ϕ coefficient from the 2×2 table of n-gram presence by edit type).

N Text Quality: Additional Results

This appendix section reports summary statistics for the text-quality and readability measures used in the analysis in Table A21, presents a within-Wikipedia placebo in Table A22, and reports the corresponding results for edited but not heavily edited articles in Table A23.

Table A21: Summary statistics for text-quality measures in heavily edited articles

	Wikipedia		Ruwiki		Mean difference	Articles
Outcome	Mean	SD	Mean	SD	R-W	
Panel A: Without contributor-risk notice						
Number of words	781.86	1,345.86	598.95	1,086.17	-182.92	52,701
Words per sentence	13.47	6.87	13.18	6.75	-0.30	52,701
Average word length	6.32	0.52	6.36	0.56	0.04	52,701
RU Flesch	30.63	17.29	30.19	17.90	-0.45	52,701
MATTR-50 $\times 100$	84.76	5.71	84.66	6.00	-0.10	52,701
Panel B: With contributor-risk notice						
Number of words	818.06	1,604.31	506.34	889.91	-311.72	1,348
Words per sentence	10.94	5.77	12.95	3.81	2.02	1,348
Average word length	6.50	0.41	6.69	0.37	0.19	1,348
RU Flesch	31.38	17.23	22.21	11.28	-9.17	1,348
MATTR-50 $\times 100$	87.00	6.25	81.67	4.74	-5.33	1,348

Notes: The sample consists of heavily edited articles observed in both Wikipedia and Ruwiki. Panel A contains articles without a contributor-risk notice; Panel B contains articles with such a notice. Means and standard deviations are computed separately across the Wikipedia and Ruwiki versions. Mean difference is the average article-level difference, Ruwiki minus Wikipedia.

Table A22 compares the May 2023 and January 2025 Wikipedia versions of the same heavily edited articles used in our baseline analysis. The changes within Wikipedia are small and mixed in direction, and they are far smaller than the corresponding Ruwiki–Wikipedia differences. For contributor-risk articles, combining the main and interaction effects, Wikipedia versions gained 63 words between May 2023 and January 2025, sentence length increased by 0.18 words, average word length fell by 0.02 characters, the Flesch score rose by 0.12 points, and MATTR-50 fell by 0.10 points. By contrast, relative to their Wikipedia counterparts, the corresponding Ruwiki versions of the same articles in 2025 contain 312 fewer words, have sentences that are 2.02 words longer and words that are 0.19 characters longer on average, and exhibit declines of 9.17 points in the Flesch score and 5.33 points in MATTR-50, as reported in Table 3 in the main text. In absolute value, these cross-platform differences are approximately five times larger for article length and between 10 and 75 times larger across the four text-quality outcomes. Moreover, the within-Wikipedia changes do not consistently move in the direction that would be required to explain the baseline results through improvements in Wikipedia. Thus, the baseline differences cannot plausibly be attributed to improvements in Wikipedia after the fork; they instead support the interpretation that text quality deteriorates following Ruwiki intervention.

We next examine the effects of Ruwiki editing on text quality among edited but not heavily edited articles. Contributor-risk notices are rare but present in both the heavily edited and non-heavily-edited samples. They appear on 2.5% of heavily edited articles, compared with only 0.5% of edited but not heavily edited articles. In some cases, relatively small changes in absolute volume may be sufficient for

Ruwiki editors to achieve their objectives. Moreover, some articles are too short to meet our threshold for heavy editing, even when the proportional intervention is substantial.

As shown in Table A23, the effects of Ruwiki editing on readability are extremely small among edited but not heavily edited articles. This further motivates our focus on heavily edited articles, for which we document a substantial deterioration in text quality in Section VI.B.3. For edited but not heavily edited articles without a contributor-risk notice, the cross-platform differences are negligible: each readability effect is approximately 0.01 standard deviations or less. Even among articles carrying a contributor-risk notice, the total differences between Ruwiki and Wikipedia remain small among non-heavily-edited articles. Ruwiki versions contain about 42 fewer words, or 0.04 standard deviations; sentence length differs by 0.03 standard deviations; average word length and the Flesch score are virtually unchanged; and MATTR-50 declines by approximately 0.07 standard deviations. The same pattern holds after controlling for article length.

Table A22: Within-Wikipedia placebo: text length and readability in the baseline heavily edited sample

	(1) *Length* Number of words	(2) *Sentence length* Words per sentence	(3) *Lexical complexity* Average word length	(4) *Readability* RU Flesch	(5) *Lexical diversity* MATTR-50 $\times 100$
Panel A: Specification without article fixed effects					
Wikipedia 2025 (vs 2023)	151.467***	0.296***	-0.029***	0.229***	-0.026
	(2.931)	(0.022)	(0.001)	(0.049)	(0.018)
Contributor-risk notice	127.151***	-2.383***	0.164***	0.832*	2.272***
	(42.094)	(0.160)	(0.012)	(0.488)	(0.174)
Wikipedia 2025 $\times$ Contributor-risk notice	-88.323***	-0.113**	0.009*	-0.106	-0.070
	(9.538)	(0.049)	(0.005)	(0.152)	(0.062)
Panel B: Specification with article fixed effects					
Wikipedia 2025 (vs 2023)	151.467***	0.296***	-0.029***	0.229***	-0.026
	(2.931)	(0.022)	(0.001)	(0.049)	(0.018)
Wikipedia 2025 $\times$ Contributor-risk notice	-88.323***	-0.113**	0.009*	-0.106	-0.070
	(9.538)	(0.049)	(0.005)	(0.152)	(0.062)
Panel C: Specification with article fixed effects and log number of words control					
Wikipedia 2025 (vs 2023)		-0.268***	0.010***	0.126	0.163***
		(0.026)	(0.003)	(0.085)	(0.023)
Wikipedia 2025 $\times$ Contributor-risk notice		0.281***	-0.018***	-0.034	-0.203***
		(0.046)	(0.005)	(0.162)	(0.063)
Wikipedia 2023 mean	616.297	13.051	6.358	30.446	84.908
Wikipedia 2023 SD	1176.832	6.366	0.562	17.705	6.055
Observations	108098	108098	108098	108098	108098
Articles	54049	54049	54049	54049	54049

Notes: This within-Wikipedia placebo uses the exact 54,049 articles in the baseline Ruwiki–Wikipedia heavily edited sample. Each article is observed in the May 1, 2023 and January 1, 2025 Wikipedia snapshots. A contributor-risk notice proxies for politically sensitive content; it warns Wikipedia editors that contributing to the article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. Panels B and C absorb article fixed effects. Panel C additionally controls for the log number of words in the corresponding Wikipedia version; its coefficient is not reported. Wikipedia 2023 means and standard deviations are calculated within the estimation sample for each outcome. Standard errors are clustered by article and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A23: Text length and readability in edited but not heavily edited articles

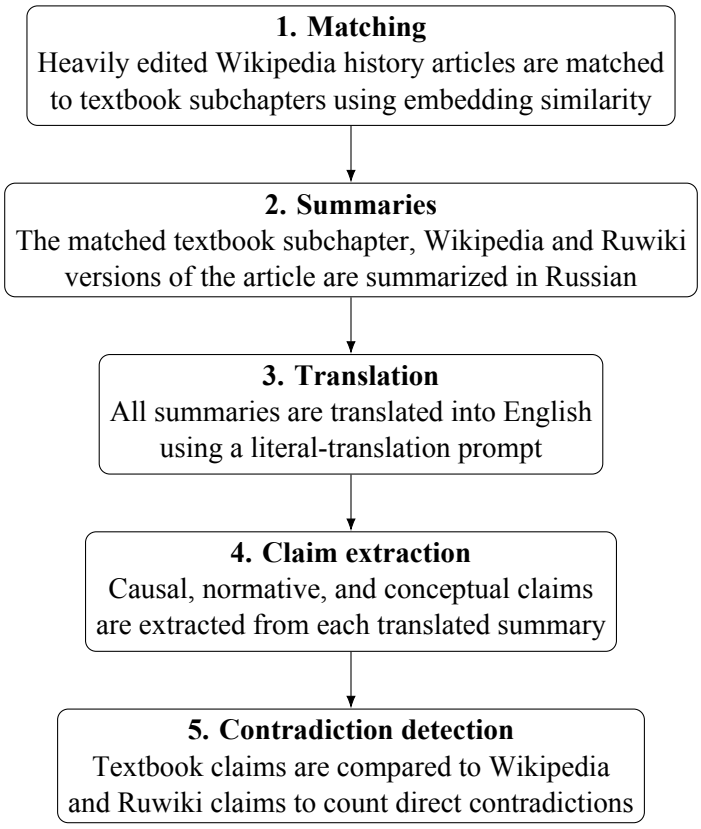
	(1) *Length* Number of words	(2) *Sentence length* Words per sentence	(3) *Lexical complexity* Average word length	(4) *Readability* RU Flesch	(5) *Lexical diversity* MATTR-50 $\times 100$
Panel A: Specification without article fixed effects					
Ruwiki (vs Wikipedia)	-6.123***	0.023***	0.007***	-0.184***	0.047***
	(0.074)	(0.002)	(0.000)	(0.005)	(0.002)
Contributor-risk notice	94.705***	0.241*	0.430***	-10.435***	-1.948***
	(22.530)	(0.145)	(0.011)	(0.368)	(0.120)
Ruwiki $\times$ Contributor-risk notice	-35.574***	0.223***	-0.012***	0.163	-0.351***
	(2.429)	(0.050)	(0.002)	(0.110)	(0.038)
Panel B: Specification with article fixed effects					
Ruwiki (vs Wikipedia)	-6.123***	0.023***	0.007***	-0.184***	0.047***
	(0.074)	(0.002)	(0.000)	(0.005)	(0.002)
Ruwiki $\times$ Contributor-risk notice	-35.574***	0.223***	-0.012***	0.163	-0.351***
	(2.429)	(0.050)	(0.002)	(0.110)	(0.038)
Panel C: Specification with article fixed effects and log number of words control					
Ruwiki (vs Wikipedia)		0.055***	0.005***	-0.202***	0.046***
		(0.003)	(0.000)	(0.010)	(0.002)
Ruwiki $\times$ Contributor-risk notice		0.302***	-0.016***	0.116	-0.355***
		(0.048)	(0.002)	(0.111)	(0.039)
Wikipedia mean	612.360	13.202	6.292	32.156	84.282
Wikipedia SD	1056.439	8.123	0.549	18.238	5.723
Observations	936172	936172	936168	936172	936168
Articles	468086	468086	468084	468086	468084

Notes: The sample consists of edited but not heavily edited articles observed in both Wikipedia and Ruwiki. A contributor-risk notice warns Wikipedia editors that contributing to the article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. Panels B and C absorb article fixed effects. Panel C additionally controls for the log number of words in the source-version text; its coefficient is not reported. Wikipedia mean and SD are calculated within the estimation sample for each outcome and are reported once at the bottom of the table. Standard errors are clustered by article and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

O Textbook Comparison Pipeline: Matching, Prompts, Examples, and Additional Results

This appendix documents the textbook-comparison pipeline used in Section VII. The goal of the exercise is to compare Wikipedia and Ruwiki articles to an external benchmark: Russia's official history textbooks used in the last year of high school. The analysis is restricted to the two 2023 Grade-11 textbooks covering Russian and world history from the end of World War II to the present: Medinsky and Torkunov (2023) and Medinsky and Chubar'yan (2023). The pipeline proceeds in five stages, summarized in Figure A19.

Figure A19: Pipeline for comparing encyclopedia articles to textbook narratives



Stage 1: Matching encyclopedia articles to textbook subchapters. We OCR-digitize both textbooks and divide them into substantive subchapters. We then embed every textbook page and every heavily edited history article using OpenAI's `text-embedding-3-large` model. Page-level textbook embeddings are aggregated to subchapter-level centroids by averaging across pages within each subchapter. For each heavily edited Wikipedia article, we identify the textbook subchapter with the highest cosine similarity and retain the 100 strongest article–subchapter matches. Manual inspection confirms that the resulting matches are substantively coherent. For example, the article on Perestroika is matched to the textbook subchapter on the reform of the Soviet political system and its consequences.

The matching step uses the Wikipedia version of the article, not the Ruwiki version. Thus, the selection of article–subchapter pairs is not mechanically driven by similarity between Ruwiki and the textbook. If anything, this procedure makes it harder to find that Ruwiki is closer to the textbook benchmark, since the selected encyclopedia articles are those whose Wikipedia versions are most similar to the relevant textbook subchapters.

Stage 2: Summarizing the matched texts. For each matched article–subchapter pair, we construct comparable summaries of three texts: the textbook subchapter, the Wikipedia version of the article, and the corresponding Ruwiki version of the article. GPT-5.5 with high reasoning effort produces a concise Russian-language summary of each text, preserving chronology, terminology, and framing. We use summaries rather than full texts in the subsequent stages to focus the comparison on the central historical narrative and to place the textbook and encyclopedia texts in a common format. All LLM prompts used at every stage of this pipeline are presented in Appendix Section O.1.

Stage 3: Translation. We translate each Russian-language summary into English with GPT-5.5 with high reasoning effort using a literal-translation prompt. The prompt instructs the model to stay close to the original structure, rhetoric, and terminology, and to avoid interpretation or summarization. This step makes the subsequent claim-extraction and contradiction-detection stages operate on a common English-language representation while preserving the framing of the original Russian texts.

Stage 4: Extracting claims. From each translated summary, GPT-5.5 with high reasoning effort extracts three types of substantive claims: causal, normative, and conceptual claims. Causal claims describe why an event occurred. Normative claims evaluate whether an action, event, or policy was justified or desirable. Conceptual claims define, classify, or name historical phenomena. The output of this stage is a structured list of claims for each textbook subchapter, Wikipedia article, and Ruwiki article.

Stage 5: Detecting contradictions. We then compare each encyclopedia article to its matched textbook subchapter. For each comparison, GPT-5.5 with high reasoning effort receives two structured claim lists: one from the textbook subchapter and one from the encyclopedia article. The model identifies pairs of claims that directly contradict one another in meaning or implication and returns the number of such contradictions, together with a brief explanation. We define contradictions narrowly. Differences in emphasis, detail, or salience are not counted unless the two texts make opposing claims. This distinction is important because the exercise is designed to measure incompatibility in historical interpretation, not differences in coverage.

Summary statistics are presented in Appendix Table A24.

Table A24: Summary statistics: contradictions with textbook and numbers of substantive claims

	Wikipedia					Ruwiki				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Contradiction count	100	0.250	0.557	0	2	100	0.080	0.339	0	2
Any contradiction	100	0.190	0.394	0	1	100	0.060	0.239	0	1
Encyclopedia claim count	100	11.840	5.768	3	38	100	11.420	6.076	3	32
Textbook claim count	100	21.340	7.733	8	48	100	21.340	7.733	8	48

Notes. Summary statistics. The unit of observation is an encyclopedia article. Sample: Wikipedia and Ruwiki versions of the article for top 100 Wikipedia history heavily edited articles matched by topic to the official history textbook.

O.1 Prompts Used in the Textbook Comparison Pipeline

This subsection reports the prompts used at each LLM stage. Summary, translation, claim-extraction, and contradiction detection calls use `gpt-5.5-2026-04-23` with high reasoning effort.

Textbook Summary Prompt

`RUSSIAN ORIGINAL:` Пожалуйста, кратко изложите этот текст в виде одного абзаца из 7–10 недлинных предложений. Сохраняйте стиль, хронологию, вокабуляр. Не добавляйте интерпретаций или оценок. Ответ строго на русском.

```
TRANSLATION: Please briefly summarize this text in the form of one paragraph
consisting of 7--10 short sentences. Preserve the style, chronology, and
vocabulary. Do not add interpretations or evaluations. The response must be
strictly in Russian.
```

Article Summary Prompt

`RUSSIAN ORIGINAL:` Пожалуйста, кратко изложите этот текст в виде одного абзаца из 5–7 недлинных предложений. Сохраняйте стиль, хронологию, вокабуляр. Не добавляйте интерпретаций или оценок. Ответ строго на русском.

```
TRANSLATION: Please briefly summarize this text in the form of one paragraph
consisting of 5--7 short sentences. Preserve the style, chronology, and vocabulary.
Do not add interpretations or evaluations. The response must be strictly in
Russian.
```

Translation Prompt

```
You are a precise translator. Translate Russian into English, staying as close as
possible to the original structure, rhetoric, and terminology. Preserve historical
names and dates. Avoid any interpretation or summarization.
```

Extract Claims Prompt

```
Enumerate all causal, normative, and conceptual claims from the text.

Output only in this exact format:

causal:
• ...
normative:
• ...
conceptual:
• ...

No explanations or extra text.
```

Contradiction Detection Prompt

```
You will be provided with the lists of normative, conceptual, and causal claims
discussed in two texts.

Identify the number of claims that contradict each other across the two texts.
Count only direct contradictions: cases where the two texts make claims that oppose
or conflict with one another in meaning or implication. Do not count differences in
emphasis, detail, or salience as contradictions.

Return your response as a short JSON with two keys:
- count: integer number of contradictory claims
- explanation: 2--3 sentences explaining briefly the main source of contradiction.
```

O.2 Details of regression specification for the Textbook Comparison

The final dataset contains 200 observations: the Wikipedia and Ruwiki versions of each of the 100 matched articles.

Formally, for article pair i and source $s \in \{W, R\}$, we estimate the linear probability model

$$Y_{is} = \beta \times \mathbb{1}\{s = W\} + X'_{is}\gamma + \phi_i + \varepsilon_{is},$$

where Y_{is} is an indicator for at least one contradiction. ϕ_i denotes article-pair fixed effects and X_{is} includes controls for the number of claims extracted from each article. For the count of contradictions outcome, we estimate the analogous PPML specification. A positive coefficient on the Wikipedia indicator means that Wikipedia is farther from the textbook benchmark than the corresponding Ruwiki article.

O.3 Additional results

Given the small number of direct contradictions, we supplement the regression estimates with exact paired tests and within-pair permutation tests that remain valid in small samples. The results are presented in Table A25.

Table A25: Small-sample paired inference for contradictions with textbook

Outcome	Ruwiki mean	Wikipedia mean	Difference	Wiki > Ruwiki	Ruwiki > Wiki	p-value
Panel A. Exact paired tests						
Any contradiction	0.060	0.190	0.130	15	2	0.002
Contradiction count	0.080	0.250	0.170	15	3	0.008
Panel B. Within-pair permutation tests						
Any contradiction	0.060	0.190	0.130	–	–	0.002
Contradiction count	0.080	0.250	0.171	–	–	0.001

Notes. The table reports small-sample inference between Ruwiki and Wikipedia history articles with Grade-11 history textbook. In Panel A, p-values are from exact two-sided binomial tests using only pairs with nonzero within-pair differences. For the binary outcome, these are McNemar-style discordant-pair tests. For the count outcome, these are paired sign tests. In Panel B, the reported difference is the coefficient on Wikipedia from a matched-pair fixed-effect linear model controlling for the total claim count. The p-values are from 10000 within-pair permutations that randomly swap the Wikipedia and Ruwiki label within each matched article pair.

As a robustness check, Table A26 reconstructs the sample by selecting the 100 article–subchapter matches with the highest similarity to the Ruwiki, rather than the Wikipedia, version of each article. The resulting estimates are consistently larger than those in the baseline sample: in the preferred specifications, Wikipedia is 14.7 percentage points more likely to contain at least one contradiction and has approximately 8.3 times as many contradictions as the corresponding Ruwiki version. This confirms that selecting matches using Wikipedia is conservative with respect to our hypothesis.

Table A26: Ruwiki-matched sample: Contradictions Between Official History Textbooks and Wikipedia vs. Ruwiki Articles

	Any contradiction			Contradiction count		
	(1)	(2)	(3)	(4)	(5)	(6)
Wikipedia (vs Ruwiki)	0.150***	0.146***	0.147***	1.705***	1.674***	2.120***
	(0.039)	(0.038)	(0.038)	(0.629)	(0.627)	(0.562)
Substantive claim counts		✓	✓		✓	✓
Article-pair FEs			✓			✓
Mean DV	0.105	0.105	0.105	0.130	0.130	0.684
Observations	200	200	200	200	200	38

Notes: This table replicates the analysis reported in Table 4 using the 100 article–subchapter pairs selected on the basis of similarity to the Ruwiki, rather than the Wikipedia, version of each article. Each specification and outcome is otherwise identical to the corresponding specification in the baseline table. Each observation is an encyclopedia article in either its Wikipedia or Ruwiki version matched to a textbook subchapter. As an outcome, we count direct contradictions between the official textbook subchapter and each encyclopedia article. The "Wikipedia (vs Ruwiki)" indicator equals one for the Wikipedia article and zero for the Ruwiki article. Columns 1–3 report OLS estimates with the binary outcome (Any contradiction); columns 4–6 report PPML estimates with the count outcome. Columns 2, 3, 5, and 6 control for total numbers of substantive claims both in the article and in matched textbook subchapter. Article-pair fixed effects are included in columns 3 and 6. Standard errors clustered by matched pair are reported in parentheses. PPML fixed-effect specifications omit matched pairs with zero total contradictions. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table A27 holds the matched article–subchapter sample fixed and examines whether Wikipedia's alignment with the textbook benchmark changed between May 2023 and January 2025; the estimates reveal no meaningful temporal shift.

Table A27: Within-Wikipedia placebo: Contradictions Between Official History Textbooks and Wikipedia at the fork in 2023 and in 2025

	Any contradiction			Contradiction count		
	(1)	(2)	(3)	(4)	(5)	(6)
Wikipedia 2023 (vs 2025)	0.010	0.022	0.020	-0.147	-0.089	-0.140
	(0.033)	(0.033)	(0.032)	(0.243)	(0.250)	(0.252)
Substantive claim counts		✓	✓		✓	✓
Article-pair FEs			✓			✓
Mean DV	0.145	0.145	0.145	0.205	0.205	1.025
Observations	200	200	200	200	200	40

Notes: This table replicates the specifications reported in Table 4 for the same 100 Wikipedia-matched article–subchapter pairs, comparing the May 1, 2023 and January 1, 2025 Wikipedia versions of each article. Each specification and outcome is otherwise identical to the corresponding specification in the baseline table. Each observation is an encyclopedia article in either its Wikipedia 2023 or 2025 version. As an outcome, we count direct contradictions between the official textbook subchapter and each encyclopedia article. The "Wikipedia 2023 (vs 2025)" indicator equals one for the Wikipedia article in its 2023 version and zero for the Wikipedia article in its 2025 version. Columns 1–3 report OLS estimates with the binary outcome (Any contradiction); columns 4–6 report PPML estimates with the count outcome. Columns 2, 3, 5, and 6 control for total numbers of substantive claims both in the article and in matched textbook subchapter. Article-pair fixed effects are included in columns 3 and 6. Standard errors clustered by matched pair are reported in parentheses. PPML fixed-effect specifications omit matched pairs with zero total contradictions. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

O.4 Examples of Contradictions

Table A28 reports examples of contradictions between Wikipedia articles and the official history textbooks.

Table A28: Largest Wiki–Ruwiki Differences in Textbook Contradiction Classifications

Cos	Article	Matched textbook subchapter	Wikipedia: contradiction number and explanation		Ruwiki: contradiction number and explanation	
0.696	Russian-Ukrainian relations (Российско-украинские отношения)	Russia today. Special military operation (SMO) (Россия сегодня. Специальная военная операция (СВО))	2	The main contradictions concern the characterization of Crimea in 2014 and Russia's actions in February 2022. Text 1 frames Crimea as having returned to Russia after a referendum and the 2022 action as a Special Military Operation for protection/security, while Text 2 describes Crimea as annexed/occupied and the 2022 action as a full-scale invasion.	0	The two sets of claims use different framing—for example, "returned to Russia"/"SMO" versus "inclusion within jurisdiction"/"military actions"—but these do not directly negate one another as stated. Most overlapping points, such as the Minsk agreements, Crimea in 2014, and the September 2022 referendums and accession treaties, are broadly compatible in factual sequence despite differing tone.
0.674	Russia and NATO (Россия и НАТО)	International relations in the 1990s–2023. The crisis of Western global domination (Международные отношения в 1990-е — 2023 г. Кризис глобального доминирования Запада)	2	The contradictions are mainly in the framing of Russia's actions in Ukraine. Text 1 describes Crimea's 2014 incorporation as "reunification" and the 2022 action as a "special military operation," while Text 2 characterizes these as "annexation" and a "full-scale invasion," respectively.	0	The two sets of claims do not directly oppose each other; they mostly address different aspects of Russia-West/NATO relations or provide compatible chronology. Differences in wording and framing, such as "coup," "reunification," or "crisis around Ukraine," are not direct contradictions because text 2 does not assert the opposite characterization.
0.663	Yeltsin, Boris Nikolayevich (Ельцин, Борис Николаевич)	Political development of the Russian Federation in the 1990s (Политическое развитие Российской Федерации в 1990-е гг.)	1	The main contradiction concerns the date of adoption of the new Constitution: text 1 says it was adopted on December 12, 1993, while text 2 says it was adopted in October 1993. Other claims mostly add different details or emphasize different causal chains without directly conflicting.	0	The claims in the two texts are largely complementary: text 1 focuses on constitutional and political-system developments, while text 2 provides a biographical chronology of Yeltsin. Shared points, such as Yeltsin's 1996 re-election and his resignation on 31 December 1999 with Putin becoming acting president, are consistent rather than contradictory.

Notes. Examples of contradictions between the encyclopedias and the textbook. Each row is one matched textbook–Wikipedia–Ruwiki triple. The presented text is the full-length stored GPT-5.5 high-reasoning contradiction explanation for each source. *Cos* stands for Cosine similarity.