

DISCUSSION PAPER SERIES

DP21698

THE ANATOMY OF CENSORSHIP AND PROPAGANDA: EVIDENCE FROM RUSSIAN WIKIPEDIAS

Ekaterina Zhuravskaya, Ulrich Matter, Ruben
Durante and Vladimir Avetian

POLITICAL ECONOMY

THE ANATOMY OF CENSORSHIP AND PROPAGANDA: EVIDENCE FROM RUSSIAN WIKIPEDIAS

Ekaterina Zhuravskaya, Ulrich Matter, Ruben Durante and Vladimir Avetian

Discussion Paper DP21698

Published 04 July 2026

Submitted 30 June 2026

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Political Economy

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Ekaterina Zhuravskaya, Ulrich Matter, Ruben Durante and Vladimir Avetian

THE ANATOMY OF CENSORSHIP AND PROPAGANDA: EVIDENCE FROM RUSSIAN WIKIPEDIAS

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 Classification: P00, D83, L82, L86, D72, C55

Keywords: Censorship, Propaganda, Autocracy, Russia, Wikipedia, Text analysis, Large Language Models

Ekaterina Zhuravskaya - ekaterina.zhuravskaya@psemail.eu
Paris School of Economics (EHESS) and CEPR

Ulrich Matter - ulrich.matter@bfh.ch
Bern University of Applied Sciences and University of St. Gallen

Ruben Durante - ruben.durante@upf.edu
National University of Singapore, ICREA-UPF, IZA, CESifo and CEPR

Vladimir Avetian - vladimir.avetian@dauphine.psl.eu

Acknowledgements

The authors are listed in reverse alphabetical order. All authors contributed equally to every part of the paper. We are grateful for helpful comments from 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, 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).

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 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, 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; 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

¹Although Ruwiki’s investors have not been publicly disclosed, an independent investigation links the project to structures associated with Yuri 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

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

Putin, and Bank Rossiya as “Putin’s wallet” (*Novaya Gazeta Europe*, 2025).

²See, 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).

³Individuals 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.”

⁴After 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.

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, 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

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

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-4o, without informing it that the exercise concerns Ruwiki, Russian censorship, or propaganda, to identify the coherent narratives implied separately by the terms characteristic of deleted text and those characteristic of inserted text. 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. Three thematic pillars organize the editorial intervention. First, on *aggression, war, and foreign policy*, Ruwiki removes references to the Russian 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 as well as other Ukrainian territories and reframes them as administratively integrated Russian regions. Second, on *domestic politics and human rights*, Ruwiki censors discourse on opposition figures, most prominently Aleksei Navalny, as well as on human rights, civil liberties, and criticism of state institutions, while reinforcing the constitutional primacy of Russian state authority. Third, on *culture and traditional values*, Ruwiki suppresses content related

⁶Deleted 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.

to LGBTQIA+ rights and Western liberal norms, while promoting traditional gender roles and framing Russia as a defender of “traditional values” against what it portrays as Western, particularly European, moral decline. These narrative characterizations constitute our first main result.

A placebo that applies the same pipeline to within-Wikipedia evolution over the twenty-month window between the fork and January 2025 recovers a qualitatively different set of narratives: editorial precision, factual completeness, and routine maintenance. This placebo exercise strongly suggests that the three pillars reflect deliberate intervention within Ruwiki rather than common temporal trends in Russian-language encyclopedic content.

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 presents 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 committed 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, including terms such as *Russian invasion*, *annexation of Crimea*, *invasion of Ukraine*, and *Russo-Ukrainian war*. These phrases are replaced with official, legalistic, and administrative terminology: *special military operation*, *military actions involving Ukraine*, and Russian administrative classifications for annexed territories. Ruwiki also adds 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. 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. 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. 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 delet-

⁷For 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).

ing 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; Guriev and Treisman, 2019, 2020, 2022; 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, 2014; Lorentzen, 2014; Shadmehr and Bernhardt, 2015; 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 and, more recently, social 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 substantive setting and the identities of the two corpora: the statistical procedure uses only anonymized corpus labels to identify differentially associated phrases, and the LLM receives only the resulting phrase lists. 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, Moksiki 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

⁸Article 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).

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](#)).

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](#)).

⁹There 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](#)).

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

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 reach-

¹⁰After 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.

able 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 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

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

¹²The 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.

¹³The 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.

¹⁴Editors 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.

Comparison design. Our primary comparison contrasts each article’s content on Wikipedia and on Ruwiki on the same date, January 1, 2025, approximately 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, tracing 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.

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 ex-

¹⁵For 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.

clude 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 annotators, 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.

¹⁶Disagreements 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.

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 3.1% 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.

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 moved from Wikipedia to Ruwiki were independent of the Russian state. Russian agents attempted to manipulate Wikipedia 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 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

¹⁷Possibly, 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](#)).

¹⁸This 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.

(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. 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 varies by type of day: ordinary weekdays, weekends, and

¹⁹The 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.

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 11.1% relative to Monday.

We also examine the pattern of activity 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% at 19:00, 44.9% at 20:00, and 56.9% at 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 Russian bank 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 Are Suppressed and What Are Promoted by Ruwiki?

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 compute a χ^2 statistic measuring the strength of its association with insertions versus deletions, and rank n-grams separately by direction. This step is purely descriptive: it identifies the n-grams most diagnostic of being inserted into or deleted from Ruwiki articles, against the null of equal frequency in the two types of text.

²⁰We 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 four distinct pages and have corpus frequency between 5 and 5,000. 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 2: LLM-based narrative extraction. We pass the top 500 n-grams in each direction to GPT-4o, using the default temperature of 1, 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 Ruwiki, censorship, propaganda, or politically motivated editing. The model is told only that the lists contain terms statistically associated with text insertions and deletions in a body of Russian-language Wikipedia articles and is asked to “identify potential narratives that might be promoted or suppressed based on these patterns.” This blinding is designed to prevent the model from importing priors about Ruwiki or being led by the framing of the exercise toward a predetermined conclusion about Russian 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.

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-4o 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 under 13 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. 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-4o, with temperature set to 1, 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 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: Russia’s image vis-à-vis the rest of the world. The first pillar concerns Russia’s foreign policy and aggression, most prominently, the war in Ukraine. At the corpus level, insertion narratives emphasize Russian territorial expansion, the legal and administrative integration of contested Ukrainian territories, the creation of new administrative units, and the justification of military action under the official euphemism of a “special military operation.” Deletion narratives are the mirror image: they remove content that points to Russian aggression, wartime violence, international criticism, sanctions, and contested territorial change. The category-level results show that this pillar is especially prominent in *history and politics*, where additions reinforce territorial integration and legal formalization of Russian control, while deletions suppress references to Russian military aggression, criticism of state policy, and con-

troversial historical episodes. The combined pattern shifts the encyclopedic record away from invasion, annexation, and international condemnation, and toward a narrative of legitimate territorial reorganization and defensive military action.²¹

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, deletion narratives include the removal of references to opposition figures and activities, most prominently those related to Aleksei Navalny, as well as broader content on civil liberties, human rights, and criticism of state institutions. Insertion narratives emphasize legal and administrative categories, resistance to Western influence, and the formalization of state authority. The category-level results show that this pillar appears most clearly in *law and order*, where additions stress state legal authority, while deletions suppress references to military aggression, human-rights abuses, and opposition voices. The combined pattern shifts content about domestic politics from contestation, opposition, and rights claims toward official legal designations and state-sanctioned institutional authority.²²

Pillar III: Culture, traditional values, and the West. The third pillar concerns culture, identity, and traditional values. At the corpus level, deletion narratives include the removal of content on LGBTQIA+ issues and other contentious social topics, while insertion narratives emphasize cultural integration and resistance to Western influence. The category-level results show that this pillar is especially pronounced in *sex and gender*, where additions align the treatment of gender and sexuality with restrictive and conservative viewpoints, while deletions reduce the visibility of LGBTQIA+ individuals and rights. It also appears in *law and order*, where additions emphasize traditional family norms and restrictions on LGBTQIA+ rights, and in *culture, media, and entertainment*, where the promotion of Russian cultural identity is combined with the suppression of controversial cultural and sexual content. The combined pattern presents Russia as a defender of traditional values against Western liberalism and moral decline, making culture

²¹This interpretation draws directly on the machine-generated narratives in Tables 1 and 2. The corpus-wide insertion narratives include “Russian Territorial Expansion,” “Legal and Administrative Changes,” “New Administrative Units,” and “Special Military Operations Justification.” The corresponding deletion narratives include “Conflict De-escalation,” “Minimizing Wartime Atrocities,” “Reducing Western Criticism,” and “Historical Revisionism.” The category-level assessment for history and politics refers to “territorial integration and legitimization of Russian control in disputed regions,” framing these actions within “legal frameworks and organizational structures,” while deletion terms retract narratives detailing “Russian military aggression.”

²²As shown in Tables 1 and 2, the deletion narratives include “Silencing Opposition,” while the insertion narratives include “Legal and Administrative Changes” and “Resistance to Western Influence.” The category-level assessment for law and order identifies additions related to “Russia’s legal and constitutional frameworks,” including “the supremacy of the Russian Constitution,” while deletions suppress information related to “human rights abuses” and “opposition voices.”

an integral part of the regime’s information-control project.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 Western liberalism and moral decline, while reinforcing respect for, and obedience to, authority within both the family and the political order. The category-level analysis shows that this intervention is both coherent and targeted. The same themes recur across topic categories, but their relative weight varies with subject matter: foreign-policy narratives dominate history and politics; legal-supremacy narratives dominate law and order; and cultural-conservative narratives dominate sex and gender as well as culture, media, and entertainment. Overall, the category-level analysis shows that Ruwiki applies a coherent ideological framework across domains while adapting the relative emphasis of its narratives to the subject matter.

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 and for both insertions and deletions, 46 of the 54 finer-level narratives (85%) preserve the hierarchy: they are classified as either still present or split. Only 8 narratives (15%) are 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 relatively few new branches. Thus, the substantive conclusions do not depend on our preferred choice of seven narratives and do not appear to 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-4o temperature 1 specification, thirteen alternative configurations identify substantively similar narratives, differing primarily in wording rather than substance. The narratives generated by the alternative models and temperatures are reported in Online Appendix Ta-

²³Tables 1 and 2 report that deletion narratives include “Censorship of Contentious Issues,” with deletions focused on “sensitive topics like LGBTQ issues.” The insertion narratives include “Cultural Integration” and “Resistance to Western Influence.” In the category-level table, the sex and gender assessment states that additions align Russia’s legal and cultural stance on gender and sexuality with “restrictive and conservative viewpoints,” while deletions “concentrate on erasing achievements and representation of LGBT individuals.” The law-and-order assessment identifies additions promoting “traditional family norms,” “restrictions on LGBT rights,” and “the supremacy of the Russian Constitution,” while the culture, media, and entertainment assessment notes deletions related to “the adult film industry” and “criticisms of governmental policies.”

bles 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 natural evolution of Russian-language encyclopedic content over the same period, rather than Ruwiki’s editorial intervention? For example, they could in principle arise mechanically from the addition of articles about the war as it unfolded to Wikipedia. To test this possibility, we apply the identical pipeline to changes within Wikipedia 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. Within Wikipedia, additions emphasize “editorial precision and factual completeness,” the expansion of historical and biographical content, sports achievements, and infrastructural developments. Deletions reflect routine encyclopedia maintenance: streamlining biographies, removing outdated technical references, and pruning unverifiable claims. There is no analogous pattern of political censorship or propaganda: no systematic removal of *Russian invasion*, legitimization of territorial claims, suppression of LGBTQIA-related content, or promotion of traditional-values narratives.

The within-Wikipedia placebo therefore rules out the main non-political alternative explanation: that the three-pillar structure reflects common temporal evolution in Russian-language encyclopedic content. The placebo instead recovers routine editorial maintenance, supporting the interpretation that the three-pillar structure reflects Ruwiki’s editorial intervention. It also addresses a methodological concern about LLM-based narrative extraction. If the LLM were biased toward producing politicized narratives whenever it encountered Russian-language text, the placebo would mirror the main analysis. Instead, it recovers the editorial-precision and routine-maintenance pattern characteristic of organic Wikipedia evolution. Overall, the placebo shows that our pipeline is sensitive to systematic editorial-intervention signals when they are present, but does not mechanically produce them when they are absent, and that the baseline comparison reflects Ruwiki’s editorial intervention rather than ordinary Wikipedia updating. We further show that the within-Ruwiki comparison over time yields patterns very similar to those from the January 2025 cross-sectional comparison between Ruwiki and Wikipedia, consistent with active Ruwiki editing driving the narrative divergence between the two platforms over the 20 months following the fork (Online Appendix Table A16 vs. 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-namespaces 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 geopo-

litical 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.

²⁴For 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.

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 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 n-grams selected with nonzero coefficients in at least 5% of iterations. Details of the empirical methodology and the cross-validation procedure used to select the penalty parameter λ are presented in

Online Appendix [H](#).

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 *LGBT*. 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 plus categories, with and without unigrams, as documented in Online Appendix [I](#).

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

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.

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

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 those 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 label 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 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 article fixed effects, so the selected n-grams identify phrases that predict the *direction* of editorial change within the same article, rather than differences in the topics or editing intensity of different articles. The L1 penalty is applied to the n-gram coefficients and selected by 10-fold cross-validation. Online Appendix K.1 reports the full specification.

As above, we use the within-Wikipedia placebo evolution between the fork in May 2023 and January 2025 to distinguish active Ruwiki editing from ordinary temporal change. Online Appendix K.2 uses our conceptual framework to guide the interpretation of the results of the baseline and placebo comparisons. It

²⁶Appendix 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*.

²⁷The article count is below the 54,049 heavily edited articles because the document-term matrix used by the LASSO requires that an article 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.

separates selected LASSO terms into three groups with distinct interpretations. Terms selected only in the Ruwiki–Wikipedia baseline comparison are the most direct markers of active divergence, because they separate the two corpora at the same date. Terms selected in both the baseline and placebo comparisons are consistent with Ruwiki staleness or passive censorship, when they appear with the opposite sign: Wikipedia changed after the fork, and Ruwiki did not fully incorporate those changes. They are consistent with different focus of the two platforms on new world events if appear with the same sign. Terms selected only in the placebo comparison characterize ordinary evolution of Wikipedia content. We therefore report the LASSO terms in three groups: baseline-only terms, terms appearing in both comparisons, and placebo-only terms.

First, we focus on terms selected by the LASSO in the baseline Ruwiki–Wikipedia comparison, but not in the within-Wikipedia placebo. They are the most diagnostic of Ruwiki-specific active editorial intervention. Panel A of Figure 5 reports the LASSO coefficients for the full corpus of heavily edited articles. The topics most systematically deleted from Ruwiki articles are dominated by the war in Ukraine and contested contemporary territorial status.²⁸ The topics most systematically inserted are also mainly geopolitical. They replace the language of invasion, Ukrainian sovereignty, and contested territorial status with official, legal, and administrative terminology.²⁹ One characteristic phrase predicting insertion across several sensitive domains is *prohibited by the Russian Federation*, a broad state legal designation used to mark organizations, materials, and activities as unlawful. Thus, references to Russian aggression and Ukrainian sovereignty are removed, while regime-sanctioned terminology for the war, restricted organizations, and territorial administration is inserted. These are precisely the kinds of words and phrases that Russian officials have demanded Wikipedia change (Fadeyev, 2023).

Panels B–D re-estimate the same LASSO logit separately within three heavily targeted LLM categories: history and politics, law and order, and sex and gender, where the distinctive pattern of Ruwiki intervention is most visible. In *history and politics* (Panel B), the pattern closely mirrors the full-corpus result. In *law and order* (Panel C), the war-related pattern remains visible but is combined with domestic politics and culture. In the cultural domain, terms such as *intersex rights* are replaced by terms such as *intersex registration*. In domestic politics, terms such as *Novaya Gazeta*, Russia’s pre-eminent independent newspaper whose editor-in-chief received the Nobel Peace Prize, are removed, while *the list of foreign*

²⁸They are associated with terms like *Russian invasion*, *annexation of Crimea*, *invasion of Ukraine*, *Russo-Ukrainian war*, *Russian-led invasion*, and *Luhansk oblast*, an occupied province, in its Ukrainian-administrative language.

²⁹The terms associated with these topics include *special military operation*, the Kremlin’s official designation for the war in Ukraine; *military actions involving Ukraine*, used instead of invasion; *oblast territory of Ukraine*, reflecting Russian administrative classifications applied to annexed territories; and *continue consider territory*.

agents, an official designation used to restrict members of the opposition and civil society, is inserted. The appearance of *Kosovo crisis* and *many countries of the world* among the inserted terms suggests an additional strategy of relativization: Ruwiki frames Russia’s actions in Ukraine by invoking other contested international cases. Thus, in law-and-order articles, Ruwiki’s edits combine 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. Deletions remove the language of rights, recognition, and protected identities, while insertions introduce the language of pathology, regulation, and prohibition. The substitutions are direct: *intersex rights* becomes *intersex registration*, and *recognition of same-sex relationships* becomes *LGBT propaganda*. This is the category in which Ruwiki’s traditional-values narrative is most explicit. Online Appendix Figure A13 provides results for other large LLM categories: culture/media/entertainment, people/person, geography, and technology, the pattern is broadly similar.

Within-Wikipedia placebo, phrase-level editing. The Wikipedia placebo recovers a qualitatively different pattern from the baseline results. We report the results for the terms that appear in LASSO only for the within-Wikipedia heavily edited articles comparison and not for the comparison between Ruwiki and Wikipedia (Online Appendix Figures A14 and A15), as well as terms predicting editing in both comparisons at the same time (Online Appendix Table A19). It is evident that there is no particular pattern here: these edits concern content that genuinely emerged during this period and routine cleanup.

Consistent with this interpretation, 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 A16).³⁰

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.

³⁰As 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 A20) and the terms that are most characteristic of the within-Wikipedia, our placebo comparison (Appendix Table A21). 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).

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; Osborneva, 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 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 inter-

action, 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.

In Online Appendix Section N, we replicate the 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.

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, we use an LLM to extract substantive causal, normative, and conceptual claims. 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. 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.

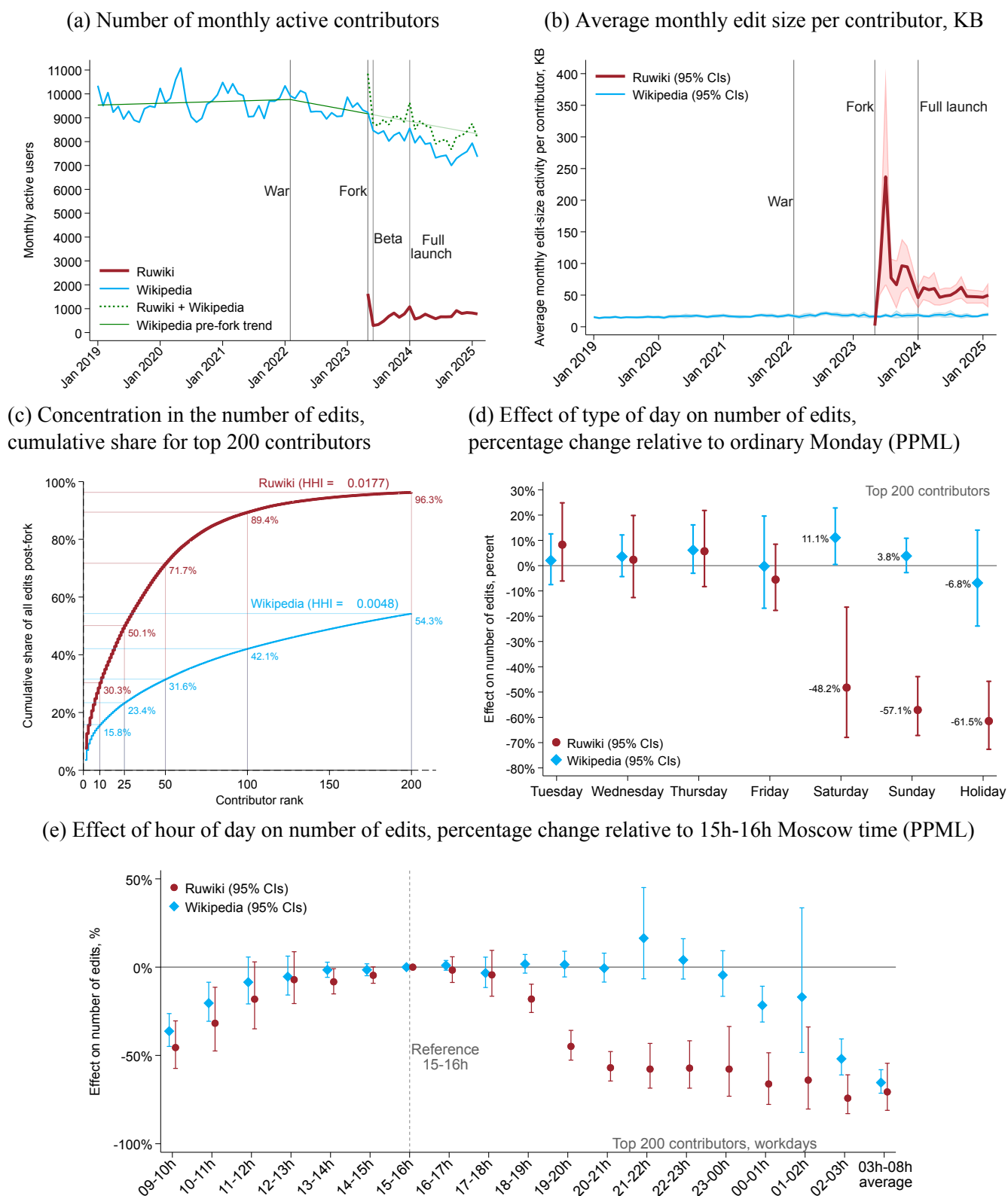
References

- 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 Moksiki, 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.
- 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, Daya Guo, Qihao Zhu, Dejian Wu, et al. 2024. "DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model." Accessed via OpenRouter API as deepseek/deepseek-chat.
- 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.
- 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.
- 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.

- Guriey, Sergei, and Daniel Treisman.** 2019. “Informational Autocrats.” *Journal of Economic Perspectives*, 33(4): 100–127.
- Guriey, Sergei, and Daniel Treisman.** 2020. “A Theory of Informational Autocracy.” *Journal of Public Economics*, 186: 104158.
- Guriey, 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 пометать “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://novyagazeta.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.
- 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.
- 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_\(historical\)](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



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 changes to the length of the article as a result of the edit. Panel C presents the cumulative share of all edits made post fork by top contributors. Panels D and E present the results of the PPML regressions which determine the dynamics of the number of edits contributors make throughout the week and during holidays as well as throughout the working day. Regressions include contributor FEs. The coefficients expressed as percentage change.

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

All text in the three panels below is reproduced verbatim from GPT-4o and has not been edited by the authors.

Overall Assessment

The addition patterns suggest a promotion of narratives surrounding Russian territorial expansion, legal integration of new areas, and cultural assimilation. There’s a strong focus on legitimizing territorial claims and portraying military actions under specific political contexts. Conversely, deletions reflect attempts to downplay Russian aggression in Ukraine, silence opposition, and minimize content related to controversial socio-political issues. This content curation appears to be part of a broader strategy to control the narrative around Russia’s geopolitical moves and domestic policy shifts, particularly in the context of international scrutiny and internal dissent.

Deletion Narratives

- 1. Conflict De-escalation:** Deletions related to involvement in the Ukraine war and related incidents suggest an attempt to downplay Russian aggression.
- 2. Silencing Opposition:** Removing references to figures like Alexei Navalny and opposition activities indicate efforts to suppress dissenting voices.
- 3. Censorship of Contentious Issues:** Some deletions focus on sensitive topics like LGBTQ issues, indicating potential censorship efforts.
- 4. Minimizing Wartime Atrocities:** Removal of content related to wartime violence and atrocities may aim to sanitize the narrative about the conflict.
- 5. Reducing Western Criticism:** Deletions about international criticism and sanctions may try to mitigate negative perceptions from Western entities.
- 6. Historical Revisionism:** Discussions about history and territorial changes are removed, potentially to control the narrative on past events.
- 7. Realignment of National Identity:** Deletions might aim to redefine the national identity by distancing from former administrative and cultural affiliations.

Insertion Narratives

- 1. Russian Territorial Expansion:** Additions focus on Russia’s inclusion of regions like the Donetsk and Luhansk People’s Republics and areas such as Zaporizhia and Kherson, underlining narratives about territorial claims and expansions.
- 2. Legal and Administrative Changes:** Content suggests the integration of territories into Russian administrative structures and legal systems, indicative of formalization of control.
- 3. Ukraine Conflict Continuation:** Narratives emphasize ongoing disputes with Ukraine, illustrating continued claims over territories and prolonged conflict.
- 4. Special Military Operations Justification:** This narrative explains Russian military actions in Ukraine as necessary operations, framing them as responses to specific territorial claims.
- 5. New Administrative Units:** New descriptions of administrative divisions and entities indicate organizational changes within recently claimed territories.
- 6. Cultural Integration:** These additions highlight efforts to integrate culturally distinct regions into the Russian sphere, possibly to promote unity and legitimacy.
- 7. Resistance to Western Influence:** Content underscores resistance to Western influence, promoting narratives against international criticism and political opposition.

Notes: This table reproduces verbatim the output of the GPT-4o narrative-extraction pipeline with temperature set to 1. 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, comprising approximately 22.4 million deleted words and 11.4 million inserted words. The LLM identified 7 distinct narratives for deletions and 7 for insertions based on statistically significant term associations (bigrams and trigrams only). The overall assessment synthesizes patterns across all narratives. Narratives are generated from the top 500 terms by χ^2 statistic in each direction.

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

All text in the table below is reproduced verbatim from GPT-4o and has not been edited by the authors.

History & Politics (N = 27,134): The addition terms reflect an interest in reinforcing narratives focusing on territorial integration and legitimization of Russian control in disputed regions. There is a significant emphasis on framing these actions within legal frameworks and organizational structures. Conversely, deletion terms disclose attempts to retract narratives detailing Russian military aggression, criticism of internal politics, and controversial issues like LGBTQ+ rights. Together, these patterns suggest a strategic intent to reinforce positive state narratives while minimizing exposure to critical perspectives, creating a controlled and favorable depiction of contemporary Russian political actions.

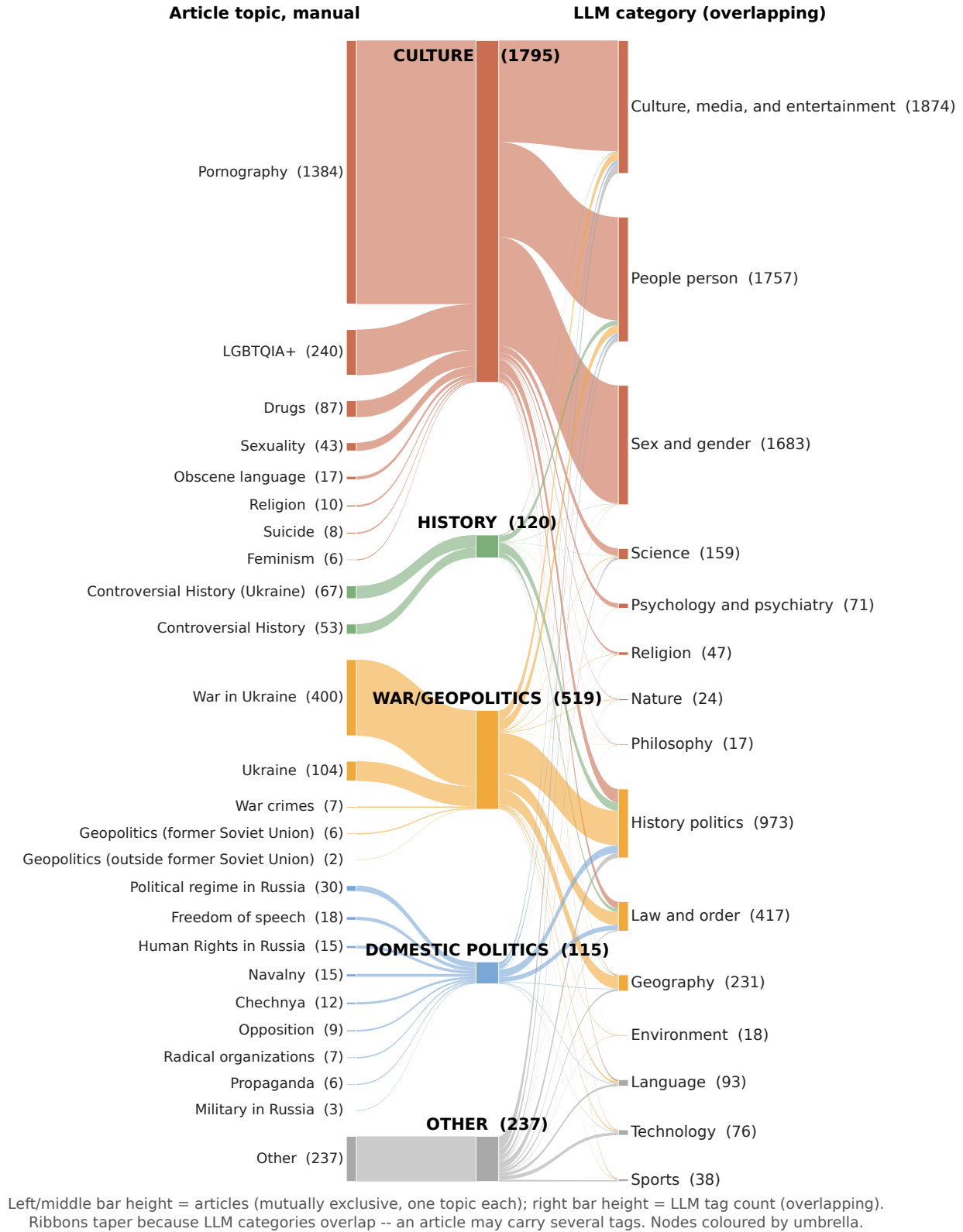
Culture, Media & Entertainment (N = 19,470): The analysis shows a dual approach in Russian media representation, where specific narratives are emphasized while others are suppressed. Additions highlight an interest in promoting Russian cultural and intellectual outputs, alongside addressing contemporary issues such as the Ukraine conflict and LGBTQ+ discussions. Meanwhile, deletions focus on downplaying controversial topics like Russian-Ukrainian aggression, the adult film industry, and criticisms of governmental policies. This reflects a broader attempt to shape public perception both domestically and internationally, presenting a curated view of Russian culture and its role on the world stage.

Law & Order (N = 3,362): The patterns in content additions and deletions in the ‘law_and_order’ category suggest orchestrated efforts to control narratives regarding Russia’s legal and constitutional frameworks. Additions focus on promoting traditional family norms, restrictions on LGBT rights, and the supremacy of the Russian Constitution. Deletions indicate suppression of information related to Russian military actions, human rights abuses, and opposition voices. This dual approach seems aimed at reinforcing state-approved ideologies while minimizing exposure to external criticism and alternative viewpoints. The content management strategy appears to be aligning legal narratives with broader socio-political agendas within Russia.

Sex & Gender (N = 2,037): The addition of content is largely aimed at aligning Russia’s legal and cultural stance on gender and sexuality with restrictive and conservative viewpoints, emphasizing legal prohibitions and traditional values. Meanwhile, deletions concentrate on erasing achievements and representation of LGBT individuals, suggesting a comprehensive effort to suppress LGBT visibility and rights in public discourse. These contrasting actions align with broader geopolitical narratives of cultural conservatism and opposition to Western ideologies.

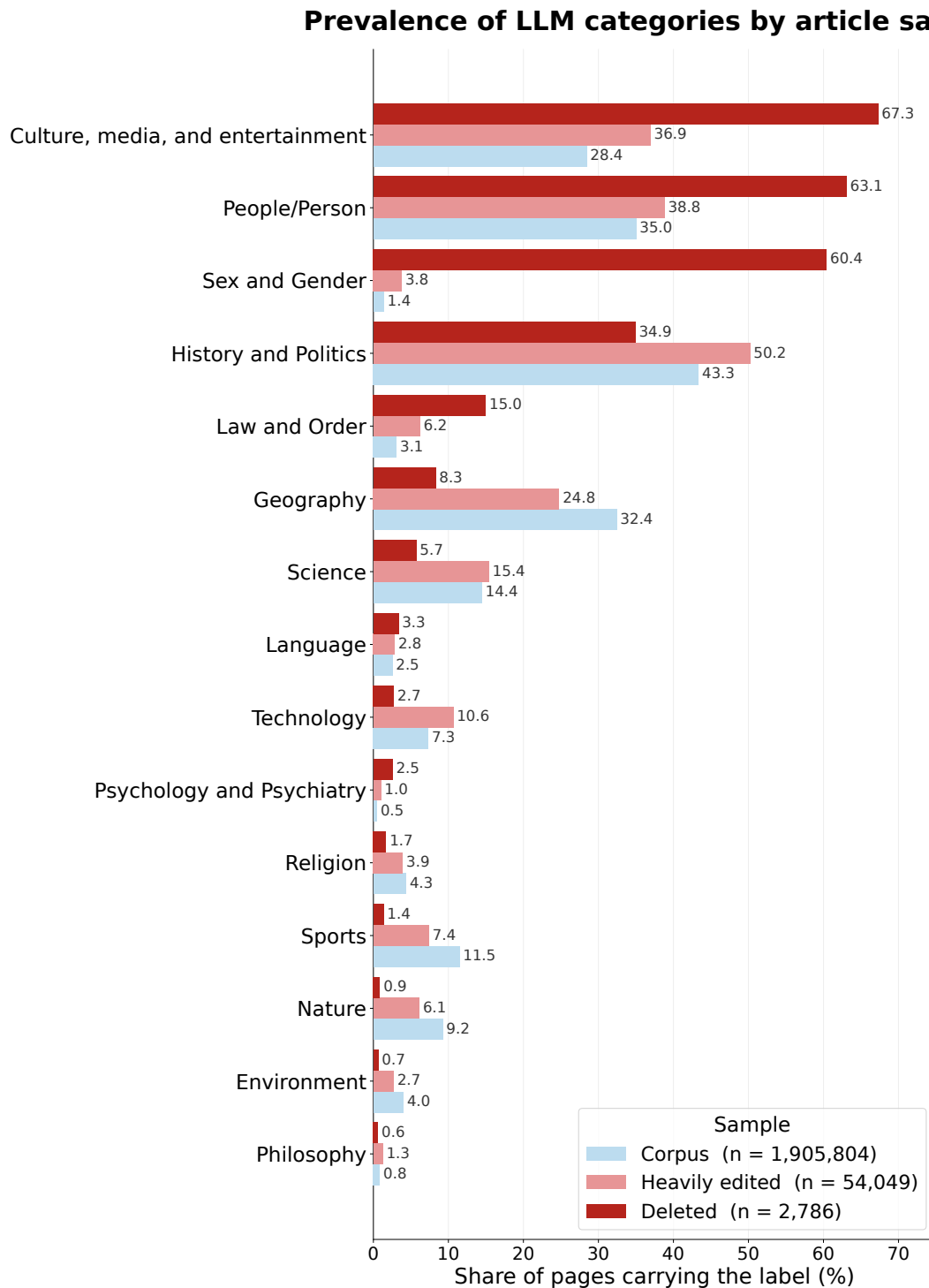
Notes: This table reproduces verbatim the output of the GPT-4o narrative-extraction pipeline with temperature set to 1. It retains the model’s original style and formulations and should not be interpreted as author-written exposition. The cross-sectional analysis compares Wikipedia to Ruwiki, identifying systematic content differences. The analysis covers 56,835 deleted and heavily edited articles, comprising approximately 22.4 million deleted words and 11.4 million inserted words. N indicates the number of heavily edited articles in each category. 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



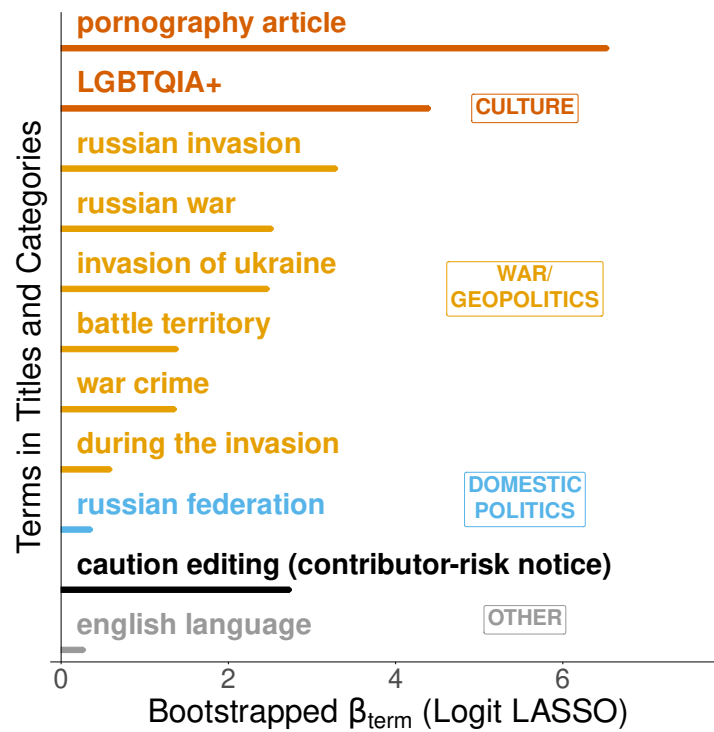
Note: The figure presents topic categorizations of all 2,786 deleted articles. It traces the flow of articles from fine-grained human coded topics, that are mutually exclusive (left) through the manually coded umbrella groups of these topics (center) to each LLM category (right). Ribbon widths proportional to the number of articles. LLM categories overlap, they are not mutually exclusive.

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



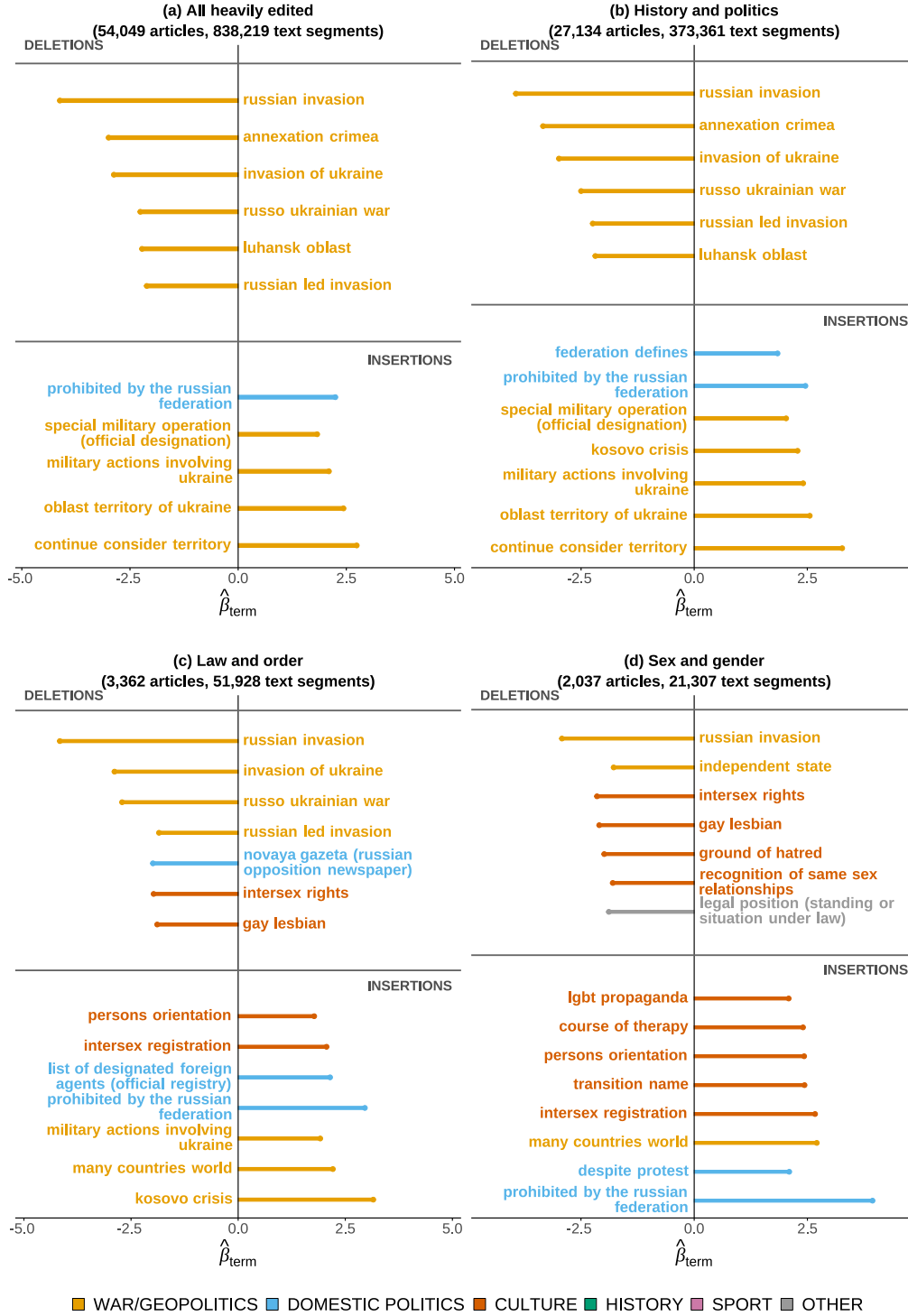
Note: The figure reports the share of articles in each sample assigned to each LLM category. Categories are not mutually exclusive: an article can be assigned to multiple categories, and therefore column percentages do not sum to 100. The full corpus consists of main-namespaces Wikipedia articles used 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



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. Positive coefficients indicate a higher probability of deletion. Terms are translated into English from the original Russian. 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. Only terms that appear as non-zero predictors in at least 5% of bootstrap samples are shown, ensuring robustness across control-group specifications.

Figure 5: LASSO predictions of deletions and insertions in heavily edited articles



Note: Estimated logit LASSO coefficients ($\hat{\beta}_{\text{term}}$) from a regularized logistic regression predicting text additions (positive) versus deletions (negative) by Ruwiki editors. Models are estimated separately by topic using lemmatized bigrams/trigrams and article fixed effects. All models use 10-fold cross-validation. Bar colors indicate semantic manually coded umbrella categories. WAR/GEOPOLITICS (gold/yellow), DOMESTIC POLITICS (blue), CULTURE (orange), HISTORY (green), SPORT (purple), OTHER (gray). (No terms in this figure are classified as HISTORY or SPORT.) Sport is added as an additional umbrella category beyond the five used for deleted articles. Terms are ordered hierarchically by the maximum absolute coefficient within umbrella categories. The unit of observation is a text segment. Sample sizes are as follows: for all heavily edited articles, 838,219 segments / 49,926 articles, for history and politics, 373,361 / 26,966; for law and order, 51,928 / 3,342; for sex and gender, 21,307 / 2,007. The article count is below 54,049 because the document–term matrix is subject to sparsity filtering before LASSO estimation.

Table 3: Text length and readability in heavily edited articles

	(1) <i>Length</i> Number of words	(2) <i>Sentence length</i> Words per sentence	(3) <i>Lexical complexity</i> Average word length	(4) <i>Readability</i> RU Flesch	(5) <i>Lexical diversity</i> MATTR-50 $\times 100$
<i>Panel A: Specification without article fixed effects</i>					
Ruwiki (vs Wikipedia)	-182.916*** (3.710)	-0.296*** (0.025)	0.039*** (0.002)	-0.447*** (0.053)	-0.100*** (0.020)
Contributor-risk notice	36.195 (44.073)	-2.537*** (0.160)	0.174*** (0.011)	0.747 (0.475)	2.240*** (0.172)
Ruwiki $\times$ Contributor-risk notice	-128.801*** (26.973)	2.313*** (0.120)	0.153*** (0.010)	-8.727*** (0.420)	-5.234*** (0.209)
<i>Panel B: Specification with article fixed effects</i>					
Ruwiki (vs Wikipedia)	-182.916*** (3.710)	-0.296*** (0.025)	0.039*** (0.002)	-0.447*** (0.053)	-0.100*** (0.020)
Ruwiki $\times$ Contributor-risk notice	-128.801*** (26.973)	2.313*** (0.120)	0.153*** (0.010)	-8.727*** (0.420)	-5.234*** (0.209)
<i>Panel C: Specification with article fixed effects and log number of words control</i>					
Ruwiki (vs Wikipedia)		0.244*** (0.030)	0.001 (0.002)	-0.294*** (0.061)	-0.214*** (0.023)
Ruwiki $\times$ Contributor-risk notice		1.982*** (0.107)	0.177*** (0.010)	-8.821*** (0.423)	-5.164*** (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*** (0.039)	0.128*** (0.040)	0.130*** (0.040)	1.139*** (0.410)	1.128*** (0.411)	1.151*** (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.³¹ 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.

³¹We 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
<i>Database & File Tracking</i>		
wiki_db	Database name (e.g., “ruwiki”)	Tracking
source_file	Original dump filename	Tracking
file_identifier	File partition identifier (e.g., “p1”, “stub1”)	Tracking
<i>Page Metadata</i>		
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
<i>Revision Metadata</i>		
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
<i>Content Metadata</i>		
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
<i>Full Content (varies by dump type)</i>		
text	Full Wikitext content of article ^b	Content
<i>Contributor Information</i>		
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 Metzner, 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 preliminary results, we additionally exclude more than two hundred technical and domain-specific terms that carry little substantive information in this setting. The resulting cleaned and lemmatized text is used to construct the n-grams employed in the subsequent statistical analyses.

We construct document-term matrices recording the occurrence of bigrams and trigrams. To focus on content-bearing expressions, we retain only n-grams containing at least one noun. We further require each n-gram to occur at least 5 and no more than 5,000 times in the corpus and to appear in at least 4 distinct articles. These restrictions exclude both rare, potentially noisy expressions and highly frequent boilerplate while preserving terms with sufficient variation for statistical analysis. We construct separate matrices for additions 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 all 1.9 million main-namespace articles in the Russian-language Wikipedia into seventeen substantive topic categories.³² 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,
```

³²Recent evidence shows that GPT-based classifications perform well against human labels and are robust to semantically equivalent prompt formulations (Asirvatham, Moksiki and Shleifer, 2026).

```
}
```

Here is the title and categories of the article:

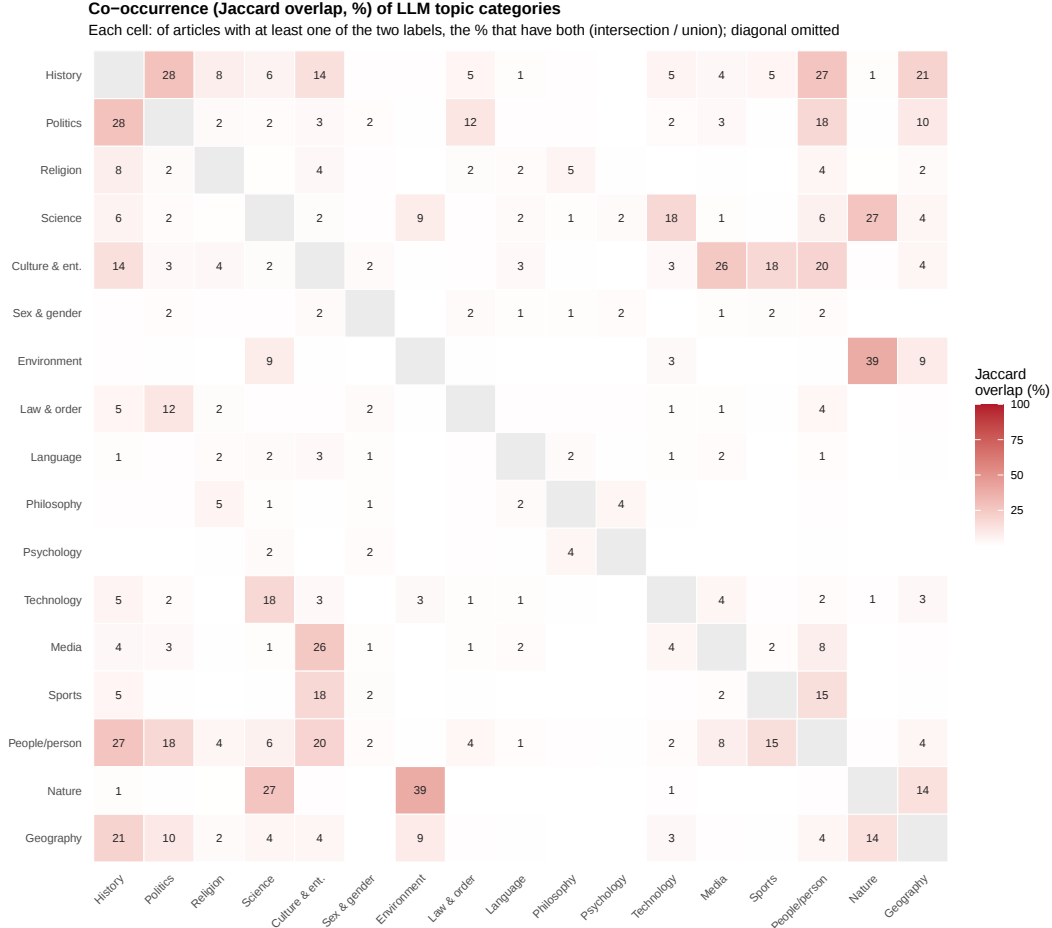
```
"[ARTICLE_TITLE -- CATEGORY_LABELS]"
```

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.

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

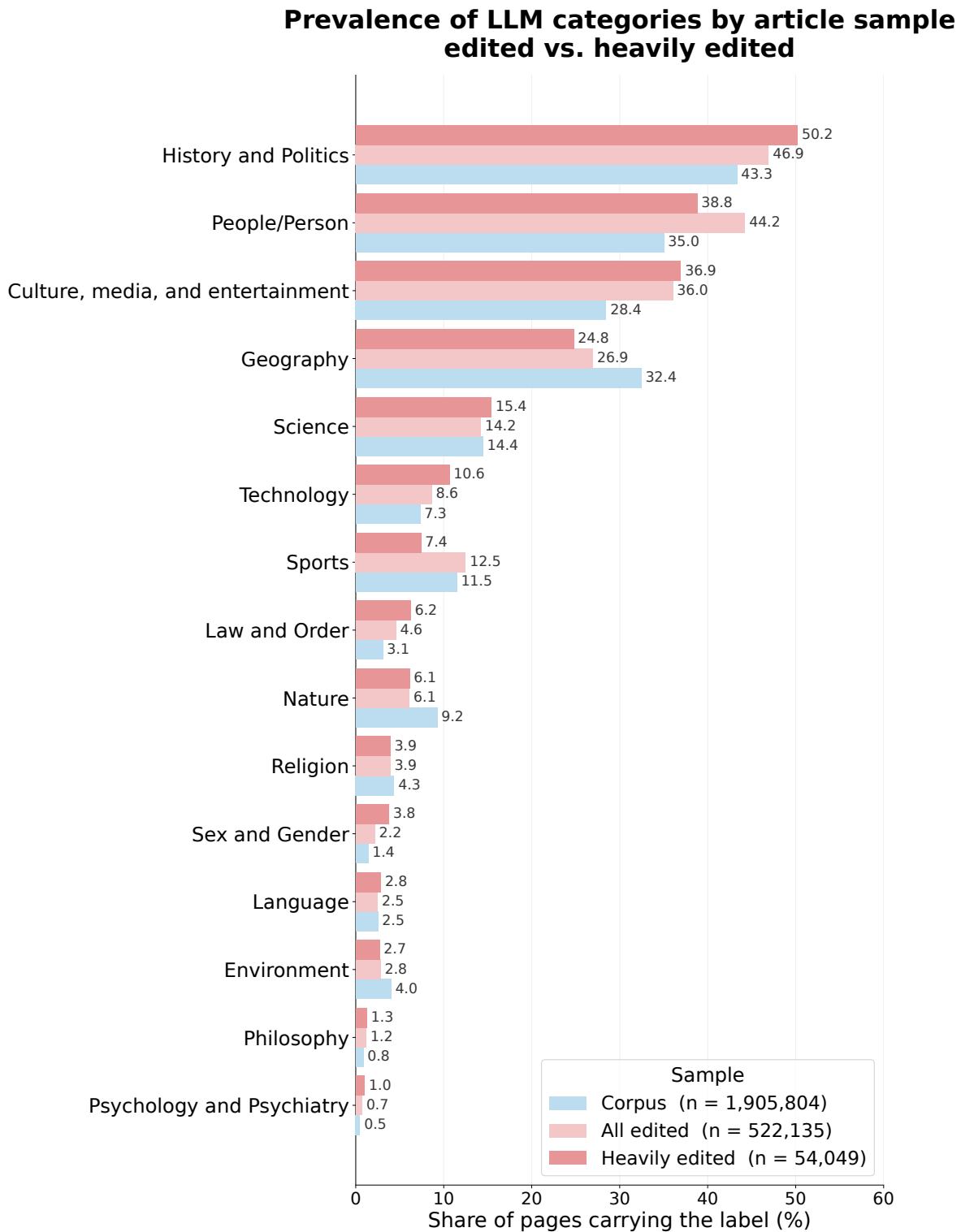


Note: We report the Jaccard overlap, in percent, for each cell, computed over the 17 binary LLM category labels; the diagonal, which corresponds to category prevalence, is omitted. To read a cell, consider all articles that have at least one of the two category labels. The reported percentage is the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

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

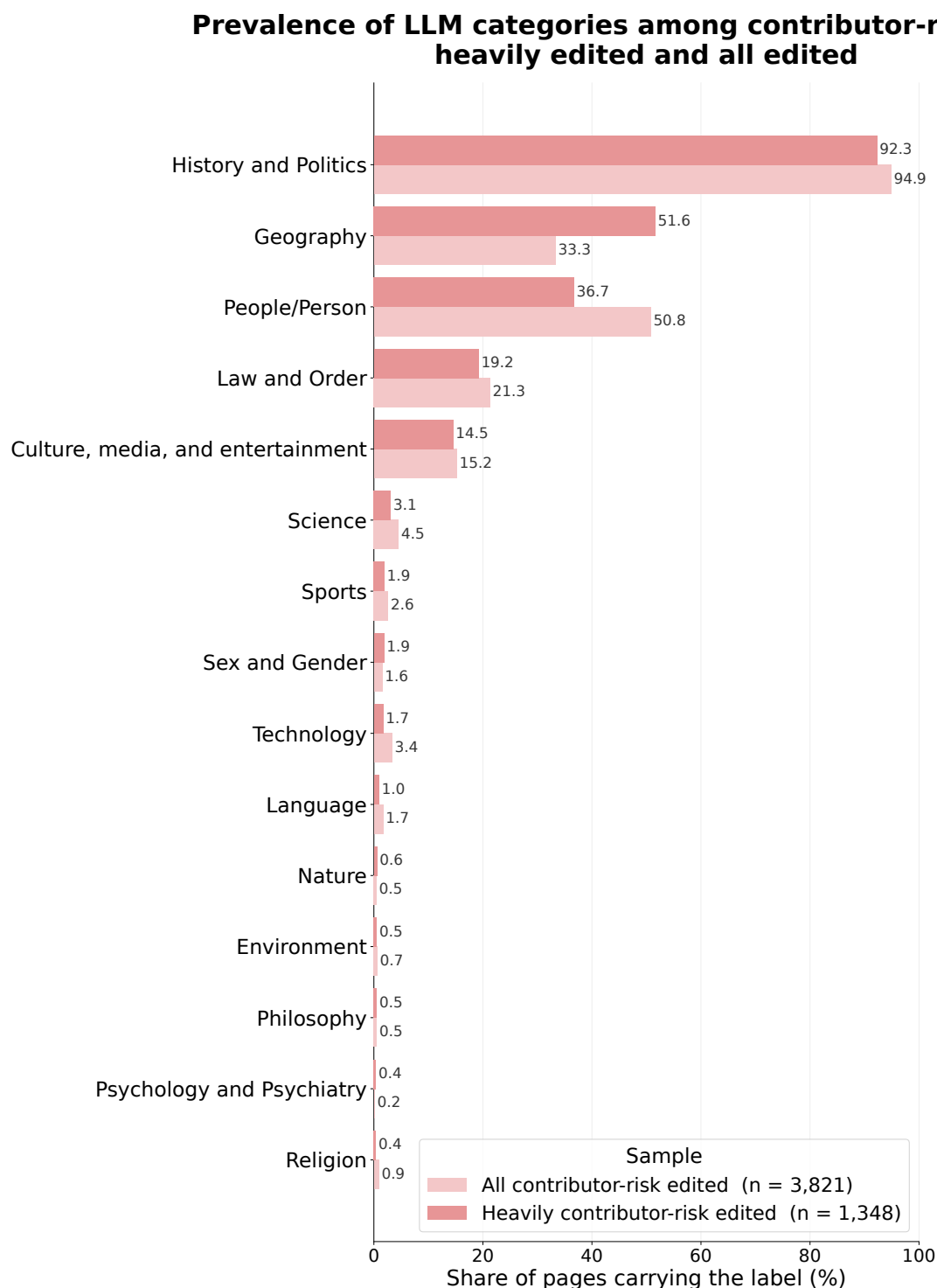
The measure ranges from 0%, when no article has both labels, to 100%, when the two labels always occur on exactly the same articles. Higher values indicate greater overlap between the two categories. Because labels are not mutually exclusive, this is an overlap measure, not a conditional probability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

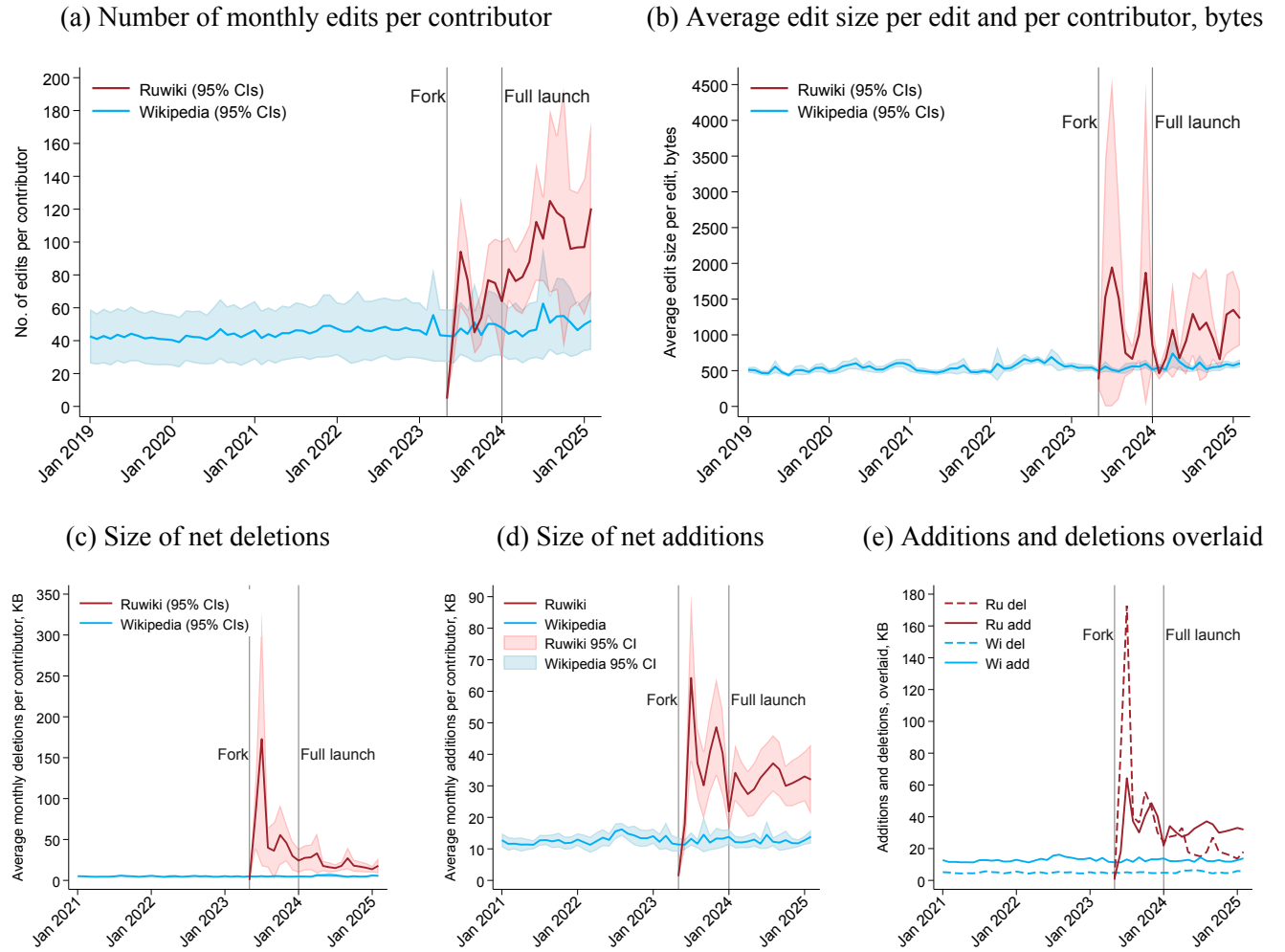


Note: The figure reports the share of articles in each sample assigned to each LLM category. Categories are not mutually exclusive: an article can be assigned to multiple categories, and therefore column percentages do not sum to 100. The samples are heavily edited articles that carry contributor-risk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

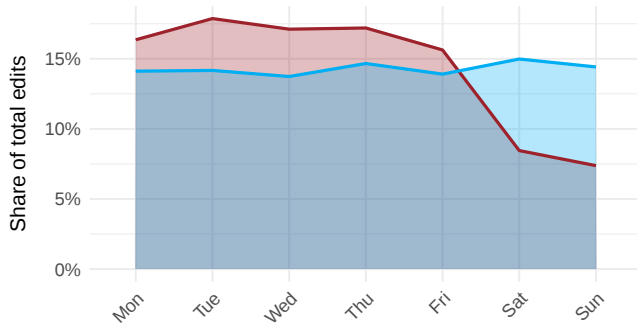
Figure A4: Averages of the number of edits, size per edit, and size of monthly additions and deletions



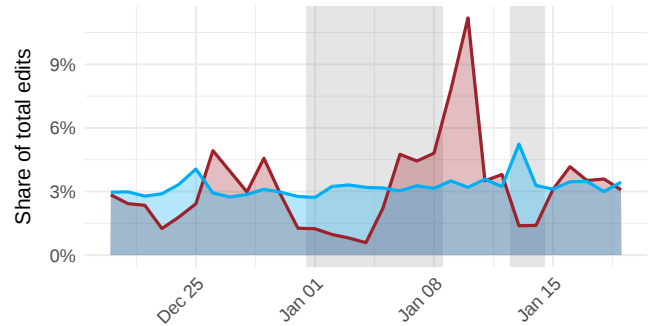
Notes: The figure is based on analysis of 1,247,135 Ruwiki edits and 7,985,350 Wikipedia edits in the post-fork period, May 2023 to February 2025, as well as 19,292,169 Wikipedia edits pre-fork starting in Jan 2019. Panel A presents the number of monthly edits per contributor. Panel B presents a proxy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

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

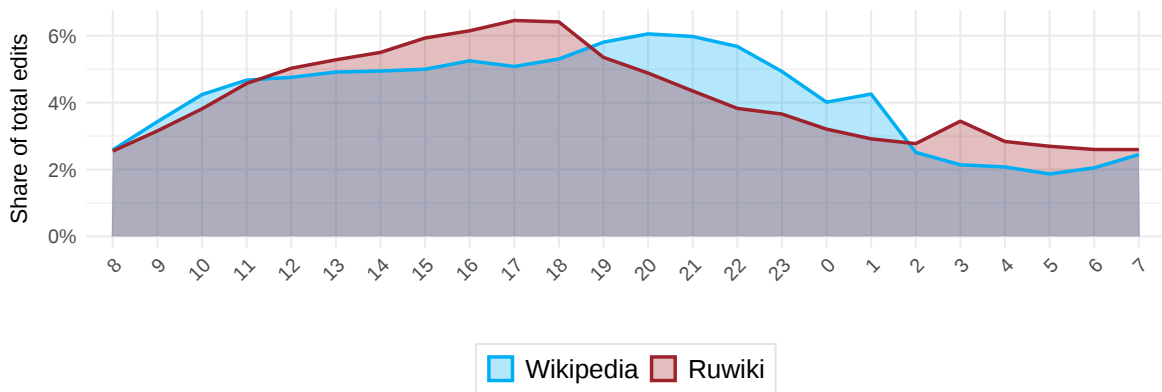
(a) Activity by day of week and holidays



(b) New Year Holidays activity



(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 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-4o with temperature 1.0) 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 that might be promoted or suppressed based on these patterns of additions and deletions. Consider the cultural, political, and social context of Russia when making your assessment.

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. For each narrative at the finer level, GPT-4o compares it with the set of narratives at the coarser level and determines whether its cross-level relationship is *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 separates; and as *new* if it has no counterpart at the coarser level. If none of these classifications applies, we treat the same narrative as persisting across the two levels.

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* relationships and unchanged 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 $\mathcal{E}$

foreach $n_h \in \mathcal{H}$ **do**

foreach $n_l \in \mathcal{L}$ **do**

$(type, w) \leftarrow \text{GPT-4o}(n_l, n_h);$

if $w > 0$ **then**

$\mathcal{E} \leftarrow \mathcal{E} \cup \langle n_l, n_h, type, w \rangle;$

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.

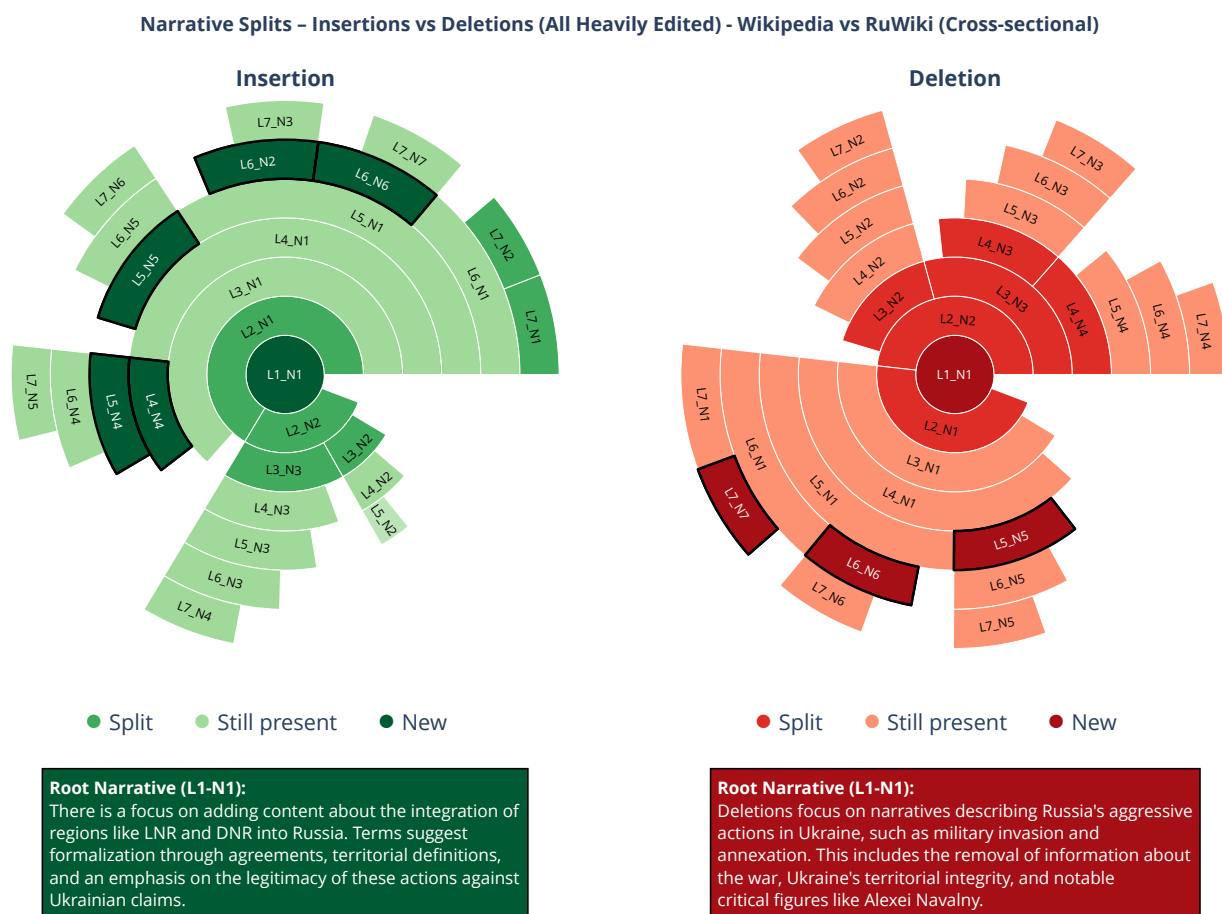
Appendix Section K.2 applies this framework to the specific tests we conduct.

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 diagram with resulting hierarchical structure of narratives in insertions and deletions based on narrative results re-running the pipeline varying the number of requested 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 most finer narratives remain nested within themes identified at coarser levels, while relatively few emerge as new branches.

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-4o. Segment size reflects the model's confidence in the relationship between levels. Colors distinguish split, still-present, and new narratives across levels, with new narratives highlighted by a bold black border. 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 via OpenRouter at temperature 1. 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](#) to [A14](#) present the results. Each configuration is identified in its table caption. All text in each of these tables is reproduced verbatim from the corresponding model and has not been edited by the authors.

Table A2: LLM Narrative Assessment (Robustness): GPT-4.1 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 in heavily edited Russian Wikipedia articles reveal a concerted effort to legitimize Russian territorial claims, normalize administrative and legal changes, and promote conservative social values, particularly regarding LGBT issues. Additions focus on legal, bureaucratic, and constitutional language to frame new territories as integral parts of Russia, while deletions systematically remove references to invasion, Ukrainian sovereignty, international sanctions, and independent media. This dual approach promotes a narrative of lawful expansion and internal order, while suppressing evidence of aggression, dissent, and human rights concerns. The result is a Wikipedia landscape that aligns closely with official Russian state narratives and minimizes critical or oppositional perspectives.

Deletion Narratives	Insertion Narratives
1. Removal of References to Russian Invasion and Aggression: Deletions systematically erase terms like 'invasion', 'aggression', and 'occupation', minimizing the perception of Russia as an aggressor in Ukraine and reframing the conflict as less severe or justified. 2. Suppression of Ukrainian Sovereignty and Resistance: References to Ukrainian territorial integrity, resistance, and leadership (e.g., Zelensky) are removed, undermining the legitimacy of Ukraine's claims and its international support. 3. Erasure of Western and International Sanctions: Mentions of sanctions, sanction lists, and international legal actions are deleted, reducing the visibility of Russia's international isolation and the consequences of its actions. 4. Downplaying of LGBT and Gender Diversity: Deletions target references to LGBT rights, sexual orientation, and gender diversity, erasing visibility and discussion of these topics in line with restrictive Russian policies. 5. Removal of Critical and Independent Media References: Mentions of independent or critical media outlets and journalists (e.g., Novaya Gazeta, Radio Svoboda, Alexei Navalny) are deleted, suppressing dissenting voices and alternative narratives. 6. Erasure of Historical and Cultural Ukrainian Identity: Deletions remove references to Ukrainian administrative units, language, and cultural markers, diminishing the distinctiveness of Ukrainian identity in favor of a more Russified narrative. 7. Suppression of War Crimes, Human Rights, and Civil Liberties: References to human rights violations, war crimes, and civil liberties are deleted, obscuring reports of abuses and undermining international legal scrutiny.	1. Legitimization of New Russian Territories: Additions emphasize the inclusion of new territories (e.g., Donetsk, Luhansk, Kherson, Zaporizhzhia) as official subjects of the Russian Federation, often referencing legal frameworks, referendums, and administrative changes. 2. Dispute of Ukrainian Sovereignty: Frequent references to Ukraine 'contesting' or 'continuing to consider' certain territories as its own, framing the situation as a legal or administrative dispute rather than an invasion or occupation. 3. Normalization of Russian Administrative Control: Additions focus on the establishment of Russian administrative structures, local governance, and legal codes in newly acquired or disputed regions, presenting these changes as routine bureaucratic updates. 4. Emphasis on Russian Legal and Constitutional Supremacy: References to the priority of Russian law and constitution, especially in the context of new territories, reinforce the narrative that Russian legal norms are now paramount in these regions. 5. Suppression of LGBT Rights and Propaganda: Additions highlight legal prohibitions against LGBT 'propaganda', same-sex marriage, and adoption, aligning with recent Russian legislative trends and promoting a conservative social agenda. 6. Framing Military Actions as 'Special Operations': The use of terms like 'special military operation' and avoidance of 'war' or 'invasion' aligns with official Russian rhetoric, downplaying the scale and aggression of military activities in Ukraine. 7. Integration of Local and Ethnic Administrative Units: Additions about the Chuvash Republic, Tatarstan, and other localities stress the formalization and reorganization of administrative boundaries, possibly to reinforce central control and national unity.

Notes: This robustness check uses GPT-4.1 with temperature=0 (deterministic mode) to assess sensitivity to temperature parameter.

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 indicate a concerted effort to reinforce Russian state narratives regarding the legitimacy of territorial changes, the normalization of Russian legal and administrative control, and the promotion of traditional values. Simultaneously, there is systematic suppression of language that frames Russia as an aggressor, supports Ukrainian sovereignty, or highlights human rights and LGBTQ+ issues. These patterns reflect the broader context of information control in Russia, where narratives aligning with state policy are promoted, and dissenting or internationally recognized perspectives are minimized or erased. The result is a Wikipedia landscape that increasingly mirrors official Russian discourse and reduces the visibility of alternative or critical viewpoints.

Deletion Narratives	Insertion Narratives
1. Suppression of 'Invasion' and 'Aggression' Terminology: Removal of terms like 'invasion', 'aggression', and 'occupation' in relation to Russia's actions in Ukraine, aiming to minimize framing the conflict as an act of war. 2. Erasure of Ukrainian Sovereignty and Integrity: Deletion of references to Ukrainian territorial integrity, sovereignty, and international support for Ukraine, downplaying the legitimacy of Ukraine's claims and international backing. 3. Minimization of Western and Ukrainian Political Figures: References to Ukrainian leaders (e.g., Zelensky), opposition figures (e.g., Navalny), and Western support are removed, reducing their visibility and perceived legitimacy. 4. Removal of LGBTQ+ and Gender Diversity References: Elimination of terms related to LGBTQ+ identities, rights, and sexual orientation, reflecting a suppression of information and visibility for these groups in line with state policy. 5. Downplaying of Civil Society and Independent Media: References to independent media, civil society, and dissenting voices (e.g., 'Новая газета', 'Эхо Москвы') are removed, limiting exposure to alternative perspectives. 6. Obfuscation of War Crimes and Human Rights Abuses: Terms referencing war crimes, human rights violations, and international legal proceedings are deleted, potentially to avoid association with such allegations. 7. Removal of Ukrainian Cultural and Administrative Markers: Erasure of Ukrainian place names, administrative units, and local governance structures, furthering the narrative that these areas are no longer Ukrainian.	1. Legitimization of New Russian Territories: Terms emphasize the inclusion of regions like LNR, DNR, Zaporizhzhia, and Kherson as new subjects of the Russian Federation, presenting their integration as legally and administratively determined by Russian law. 2. Delegitimization of Ukrainian Sovereignty: Phrases frame Ukrainian claims over disputed territories as contested or invalid, often stating that Ukraine 'continues to dispute' or 'claims' territory now 'defined' by Russia as its own. 3. Normalization of Russian Legal and Administrative Control: Emphasis on Russian legal codes, administrative units, and governance structures for newly integrated territories, presenting changes as routine and legally grounded. 4. Framing of Military Actions as 'Special Operations': References to the conflict in Ukraine are couched in terms like 'special military operation', aligning with official Russian terminology and avoiding the word 'war' or 'invasion'. 5. Promotion of Traditional and Conservative Values: Strong focus on anti-LGBTQ+ legislation and rhetoric, with references to bans on propaganda, same-sex marriage, and adoption, aligning with official state positions on 'traditional values'. 6. Emphasis on Russian Constitutional Supremacy: Highlighting the primacy of the Russian Constitution and its norms in the governance of new territories and in legal disputes, reinforcing a narrative of legal legitimacy. 7. De-emphasis of Ukrainian Administrative Structures: Frequent references to dissolving or reclassifying Ukrainian administrative units (e.g., oblasts, raions) into Russian structures, suggesting the erasure of Ukrainian governance.

Notes: This robustness check uses GPT-4.1 with temperature=0.5 to assess sensitivity to temperature parameter.

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

The editing patterns reveal a dual strategy: additions reinforce Russian sovereignty, the integration of newly claimed territories, and anti-LGBT/right-libertarian legal policies; at the same time, deletions suppress narratives critical of Russia's war in Ukraine, erase references to Ukrainian sovereignty, remove content on human rights and minority issues, and eliminate dissenting media voices. This mirrors official state positions on war, territorial integrity, and social values. The combination of heavy additions related to legitimizing new regions and sweeping deletions of critical terminology indicates a pronounced effort to rewrite historical, political, and social realities in line with Kremlin policy and ideology, both domestically and in externally-visible knowledge repositories like Wikipedia.

Deletion Narratives	Insertion Narratives
1. Removal of Russian Aggression and Invasion Language: Phrases directly calling Russia's actions an 'invasion' or 'aggression' (e.g., 'вторжение россия', 'российско украинский война', 'российский агрессия') are systematically deleted, reducing the visibility of narratives framing Russia as the aggressor. 2. Suppression of Ukrainian Sovereignty and Integrity: Terms upholding Ukrainian territorial integrity and independence, as well as references to 'целостность украина' and international sanctions, are removed, downplaying narratives that challenge Russian claims. 3. Removal of Western and Opposition Media References: References to independent media, dissident voices, and Western/critical sources (e.g., 'новый газета', 'эхо москва', 'радио свобода', 'алексей навальный') are prioritized for deletion. 4. Erasure of LGBT and Gender Equality Discourse: Terminology normalizing LGBT identities, sexual orientation, and gender diversity is deleted, including information on rights and visible representation, aligning with official anti-LGBT policies. 5. Downplaying of Russian Military Failures or Losses: Mentions of Russian losses, failures or negative outcomes in the war (e.g., 'протест против', 'война против', 'российский солдат', 'итог вместе клуб') are removed to avoid negative perceptions. 6. Omission of Non-Russian Historical and Local Administrative Details: References to Ukrainian place names, local councils, and detailed historical nuances reflecting non-Russian administration are targeted for deletion, fostering a Russian-centric narrative. 7. Removal of Societal Criticism and Rights-Based Language: Terms supporting human rights, criticism of state practices, and freedom of speech are deleted, leading to the suppression of liberal or critical discourses on governance and rights.	1. Legitimization of New Russian Territories: The frequent addition of phrases assigning regional status to Donetsk, Luhansk, Kherson, and Zaporizhzhia as Russian federal subjects creates a narrative legitimizing these regions as part of Russia, reflecting official state positions post-2022. 2. Emphasis on Ongoing Disputes with Ukraine: The narrative repeatedly states that Ukraine continues to dispute Russia's claims and control over contested territories, but frames the Russian perspective as authoritative. 3. Reinforcement of Russian Legal and Administrative Frameworks: Descriptions of local governance, legal changes, incorporation procedures, and official recognition emphasize the integration of new regions into Russian law, suggesting normalization and bureaucratic control. 4. Suppression of LGBT Rights and 'Propaganda': Multiple additions outline Russia's prohibitions on LGBT activities, direct references to 'LGBT propaganda', bans on same-sex marriage, adoption, and distinctions from Western practices, depicting Russia as upholding 'traditional values'. 5. Anti-Western and Anti-Sanctions Sentiment: Some entries emphasize Western challenges and Russia's resistance, referencing sanctions and the global context—portraying Russia as an embattled but sovereign nation resisting external pressure. 6. Highlighting Russian Administrative and Statistical Details: There is heavy emphasis on demography, settlement status, and administrative classifications, demonstrating control and organization—especially for new or disputed regions. 7. Legal and Political Prioritization of Russian Sovereignty: Additions focus on the supremacy of Russian law, referencing priority of the Russian constitution and legal codes over others, reinforcing the primacy of Russian governance in all incorporated areas.

Notes: This robustness check uses GPT-4.1 with temperature=1 to assess sensitivity to temperature parameter.

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 Wikipedia articles suggest a concerted effort to shape narratives around Russia's geopolitical actions, particularly regarding Ukraine. Additions emphasize the integration of disputed territories into Russia, legal reforms, and cultural narratives that support these actions. In contrast, deletions focus on removing content that highlights military aggression, opposition figures, and international sanctions. This indicates an attempt to promote a narrative of legitimacy and historical continuity while suppressing dissenting views and critical perspectives. The focus on LGBTQ+ rights in both additions and deletions suggests a complex narrative around social policies, reflecting broader cultural and political dynamics within Russia.

Deletion Narratives

- 1. Downplaying Military Aggression:** Deletions remove references to Russia's military actions in Ukraine, including terms like 'invasion' and 'aggression'.
- 2. Censorship of Opposition Figures:** Content related to opposition figures and critical media, such as Alexei Navalny and Novaya Gazeta, is being removed.
- 3. Suppression of LGBTQ+ Content:** Deletions focus on removing content related to LGBTQ+ rights and identities, including sexual orientation and gender identity.
- 4. Minimizing Ukrainian Sovereignty:** Content that emphasizes Ukraine's territorial integrity and sovereignty is being removed, including references to annexation.
- 5. Erasing Historical Context:** Deletions remove historical context related to Russian-Ukrainian relations, including past conflicts and agreements.
- 6. Reducing Visibility of Sanctions:** Content related to international sanctions against Russia is being removed, possibly to downplay their impact.
- 7. Limiting Criticism of Russian Policies:** Deletions focus on removing criticism of Russian policies, including those related to human rights and international law.

Insertion Narratives

- 1. Integration of New Territories:** Additions focus on the integration of regions like Luhansk and Donetsk into the Russian Federation, emphasizing legal and administrative changes.
- 2. Legal and Administrative Reforms:** Content highlights legal reforms and administrative changes within Russian regions, including the establishment of new territorial units.
- 3. Military Operations in Ukraine:** Additions describe ongoing military actions in Ukraine, focusing on territorial claims and military engagements.
- 4. Cultural and Demographic Changes:** Content includes demographic data and cultural shifts in newly integrated regions, reflecting changes in population and cultural identity.
- 5. LGBTQ+ Rights and Restrictions:** Additions discuss legal restrictions on LGBTQ+ rights, including bans on propaganda and same-sex marriage.
- 6. International Relations and Sanctions:** Content addresses Russia's international relations, including sanctions and diplomatic actions related to territorial disputes.
- 7. Historical and Cultural Narratives:** Additions focus on historical and cultural narratives, emphasizing Russia's historical ties to disputed territories.

Notes: This robustness check uses GPT-4o with temperature=0 (deterministic mode) to assess sensitivity to 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 additions and deletions in these Wikipedia articles suggest a concerted effort to promote narratives that support Russia's territorial claims and legal justifications for recent actions, particularly regarding Ukraine. There is a clear attempt to integrate new territories into the Russian Federation both administratively and culturally. Conversely, deletions focus on minimizing references to military aggression, opposition figures, and human rights issues, as well as obfuscating territorial disputes and international sanctions. This pattern reflects a broader strategy of narrative control, aiming to legitimize Russia's actions while suppressing dissenting viewpoints and critical historical contexts.

Deletion Narratives	Insertion Narratives
1. Minimizing Military Aggression: Deletions focus on minimizing references to Russian military aggression and the invasion of Ukraine. 2. Reducing References to Opposition Figures: References to opposition figures and organizations, such as Alexei Navalny, are being removed. 3. Downplaying Human Rights Issues: Deletions aim to reduce mentions of human rights issues, such as sexual orientation and freedom of speech. 4. Obfuscating Territorial Disputes: Mentions of territorial disputes and annexations, particularly regarding Ukraine, are being suppressed. 5. Sanctions and International Isolation: References to international sanctions and Russia's isolation on the world stage are being minimized. 6. Reducing Historical Context: Historical contexts that might negatively frame Russia's actions, such as past conflicts, are being reduced. 7. Suppressing Criticism of Government: Criticisms of the Russian government and its actions, especially related to the war, are being removed.	1. Integration of New Territories: Additions focus on the integration of new territories into the Russian Federation, particularly regions like the Donetsk and Luhansk People's Republics, and the legal frameworks supporting these actions. 2. Legitimization of Territorial Claims: There is an emphasis on legitimizing territorial claims through legal and political measures, often in opposition to Ukrainian claims. 3. Administrative Reorganization: The restructuring of administrative divisions within Russia, including the establishment of new regions and changes to existing ones, is highlighted. 4. Cultural and Historical Integration: Narratives supporting the cultural and historical integration of newly claimed territories into Russia are evident. 5. Military Operations Justification: Additions justify military actions in Ukraine by framing them as necessary for the protection and integration of Russian territories. 6. Legal and Constitutional Changes: Changes to legal and constitutional frameworks are emphasized to support the integration of new territories and ensure compliance with Russian law. 7. Opposition to Western Influence: There are narratives opposing Western influence and policies, particularly regarding social issues like LGBT rights.

Notes: This robustness check uses GPT-4o with temperature=0.5 to assess sensitivity to 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 patterns of additions and deletions in Russian-language Wikipedia articles suggest a concerted effort to shape narratives about national identity, territorial integrity, and governance. Additions are primarily focused on integrating newly claimed territories into the Russian Federation, reinforcing state authority and aligning social policies with traditional values. Conversely, deletions indicate a tendency to suppress negative depictions of Russia's international actions and internal opposition. This suggests a broader strategy to control the narrative around Russia's geopolitical moves and domestic policies, presenting a cohesive and authoritative image while downplaying controversies.

Deletion Narratives	Insertion Narratives
1. Suppression of Russian Military Actions: Efforts to suppress information on Russian military actions in Ukraine, including the narrative of invasion and occupation. 2. Minimizing International Relations Issues: Attempt to minimize discussions on international relations issues, such as sanctions and territorial integrity concerns related to Ukraine. 3. Downplaying Opposition Figures: Decreases in coverage of Russian political opposition, hinting at efforts to diminish their visibility and influence in public discourse. 4. Censorship of LGBTQ+ and Sexual Content: Removal of content related to LGBTQ+ issues and sensitive sexual content, aligning with government policies on morality and traditional values. 5. Erasing Historical and Cultural Contexts: Efforts to erase or downplay historical and cultural contexts, such as contributions and experiences relating to Ukraine and other regions. 6. Sanitizing Media and Public Information: Attempts to sanitize media coverage and limit information that could reflect negatively on Russian domestic and foreign policy. 7. Removing References to Foreign Influence: Dealt with foreign influence by removing mentions of foreign opinions, sanctions, or cultural exchanges that contrast with official Russian positions.	1. Integration of New Subjects into Russia: The narrative focuses on the integration of new regions, such as Luhansk and Donetsk, into the Russian Federation, emphasizing territorial claims and sovereignty, particularly amidst ongoing disputes with Ukraine. 2. Expanding Russian Federal Authority: Emphasizes the expansion and authority of the Russian Federation over newly claimed territories, highlighting legal and governmental measures to consolidate these regions. 3. Redefining Regional Borders: Discusses changes to regional borders within Russia, including reorganization and integration efforts, often involving territorial disputes with neighboring countries. 4. Legal and Administrative Restructuring: Details legal changes and administrative restructuring within the Russian Federation, focused on integrating new territories and establishing new governance frameworks. 5. Military Operations in Ukraine: Covers narratives surrounding Russian military operations in Ukraine, including territorial gains and strategic objectives, amidst international scrutiny. 6. Cultural and Social Integrations: Focuses on cultural and social integration efforts of new territories into Russian society, emphasizing shared values and historical connections. 7. LGBTQ+ Legislation and Propaganda: Highlights narratives around the Russian government's stance on LGBTQ+ issues, including legislation and social policies related to family values and propaganda laws.

Notes: This robustness check uses GPT-4o with temperature=1 to assess sensitivity to 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 systematic narrative manipulation serving Russian state interests. Additions legitimize territorial annexation through bureaucratic normalization, present occupation as disputed rather than illegal, and promote anti-LGBTQ+ moral panic as state ideology. The 'special military operation' euphemism replaces accurate war terminology. Deletions erase evidence of Russian aggression, international condemnation, Ukrainian sovereignty, domestic opposition, and human rights violations. Together, these patterns transform Wikipedia from neutral documentation into state-aligned propaganda. The technical flooding with administrative details obscures ideological changes. This represents coordinated information warfare: additions build alternative reality supporting Russian claims while deletions eliminate contradicting evidence. The scope and consistency suggest organized effort rather than organic editing, fundamentally compromising Wikipedia's reliability on Russia-Ukraine conflict and Russian domestic politics. The patterns align precisely with Kremlin messaging priorities across state media.

Deletion Narratives	Insertion Narratives
1. Erasure of Russian Aggression: Systematic removal of terms describing Russian military actions as 'invasion,' 'aggression,' or 'war.' Deletions eliminate references to 'Russian invasion of Ukraine,' 'Russian-Ukrainian war,' and 'aggression,' replacing factual descriptions with euphemistic alternatives. This rewrites historical record to minimize Russia's role as aggressor. 2. Suppression of Ukrainian Sovereignty: Deletions remove references to Ukrainian territorial integrity, sovereignty, and legitimate governance. Terms about 'violation of Ukraine's integrity,' 'President of Ukraine,' and Ukrainian administrative structures are eliminated, erasing Ukraine's legal status and governmental legitimacy in favor of Russian territorial claims. 3. Removal of Sanctions and International Condemnation: Extensive deletion of references to international sanctions, sanction lists (USA, Canada, EU), and diplomatic isolation. Removes evidence of global response to Russian actions, creating impression of international acceptance or irrelevance of Western opposition to Russian policies. 4. Elimination of Dissent and Opposition: Deletions remove references to opposition figures (Alexei Navalny), independent media (Novaya Gazeta, Echo of Moscow, Radio Liberty), protests, and criminal cases against dissidents. This erases evidence of internal opposition and repression, presenting false image of societal consensus. 5. Suppression of LGBTQ+ Normalization: Deletions remove neutral or positive references to LGBTQ+ individuals, same-sex couples, sexual orientation, and gender identity. Terms about rights, freedoms, and normal representation are eliminated while negative framing (in additions) remains, creating one-sided portrayal supporting discriminatory policies. 6. Removal of Ukrainian Geographic and Administrative Identity: Systematic deletion of Ukrainian place names, administrative codes (KOATUU), postal codes, and local governance structures in occupied territories. Removes Ukrainian administrative framework, facilitating replacement with Russian systems and erasing Ukrainian territorial organization. 7. Erasure of War Crimes and Human Rights Violations: Deletions remove references to violations of rights and freedoms, European Court of Human Rights, criminal proceedings, and violence. Eliminates documentation of abuses, legal accountability mechanisms, and international human rights frameworks that could document Russian actions.	1. Legitimization of Territorial Annexation: Extensive additions establish annexed Ukrainian territories (Donetsk, Luhansk, Zaporizhzhia, Kherson) as legitimate subjects of the Russian Federation. Terms emphasize legal frameworks, treaties, and administrative integration, presenting annexation as a formal constitutional process rather than occupation. References to 'new subjects,' 'composition of Russian Federation,' and 'treaties' create bureaucratic normalization of territorial seizure. 2. Disputed Sovereignty Framing: Additions systematically frame territorial control as 'disputed' rather than occupied, using phrases like 'Ukraine disputes' and 'continues to consider territory' to create false equivalence between aggressor and victim. This narrative suggests legitimate disagreement rather than violation of international law, normalizing Russia's territorial claims through linguistic balance. 3. Administrative Normalization: Detailed additions about administrative structures, municipal boundaries, and local governance in occupied territories (Chuvash Republic frameworks applied to DNR/LNR) create appearance of normal state functioning. References to laws, status changes, and territorial organization present occupation as routine administrative activity, obscuring military control. 4. Anti-LGBTQ+ Moral Panic: Extensive additions link LGBTQ+ identities with pedophilia, propaganda, and foreign agents. Terms emphasize legal prohibitions, constitutional priorities, and comparisons to restrictive regimes (DPRK, Saudi Arabia). This narrative positions Russia as defender of traditional values against Western moral corruption, justifying repressive legislation through manufactured threat. 5. Special Military Operation Euphemism: Consistent use of 'special military operation' instead of 'war' or 'invasion' reflects official state terminology. Additions frame military actions as limited, technical operations rather than full-scale invasion, minimizing severity and international law violations through bureaucratic language. 6. Constitutional Supremacy Doctrine: Additions emphasize primacy of Russian constitution over international norms, with references to constitutional priorities, federal authority, and legal frameworks. This narrative establishes domestic law as superseding international obligations, providing legal justification for actions condemned globally. 7. Technical Bureaucratic Detail Flooding: Massive additions of administrative codes, water resource classifications, territorial classifications, and municipal reorganizations create information overload. This obscures political content through technical minutiae, making controversial changes appear as routine database updates rather than ideological editing.

Notes: This robustness check uses Claude Sonnet 4.5 with temperature=0 (deterministic mode) to assess sensitivity to 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 systematic information warfare supporting Russian state narratives. Additions legitimize territorial annexation through bureaucratic normalization, frame the invasion euphemistically, promote anti-LGBTQ+ legislation, and establish constitutional isolationism. Deletions erase evidence of aggression, Ukrainian sovereignty, international condemnation, domestic opposition, and human rights discourse. Together, these create an alternate reality where annexed territories are legitimately Russian, military aggression is administrative procedure, international isolation doesn't exist, opposition is invisible, and restrictive social policies are normalized. The scale and coordination suggest organized effort to rewrite Wikipedia as historical record supporting state positions. This represents weaponization of crowd-sourced knowledge platforms for authoritarian narrative control, attempting to establish Russian government framing as factual baseline for future reference.

Deletion Narratives

1. Suppression of Invasion Terminology: Systematic removal of terms describing Russian actions as 'invasion,' 'aggression,' or 'war.' Deletions eliminate language framing events as Russian military attack on Ukraine, replacing it with neutral or euphemistic alternatives. This removes historical record of international consensus on Russian responsibility for the conflict.

2. Erasure of Ukrainian Sovereignty Markers: Removal of Ukrainian administrative codes, postal indices, telephone codes, and local council references for occupied territories. Deletions eliminate evidence of Ukrainian governmental structures and territorial integrity, erasing the administrative reality that these areas were Ukrainian-controlled before occupation.

3. Deletion of Territorial Integrity References: Removal of language about Ukraine's territorial integrity, sovereignty violations, and international recognition of borders. Deletions eliminate context establishing Ukrainian legal ownership of occupied territories and international condemnation of annexation, supporting narrative that territorial changes are legitimate.

4. Suppression of Opposition Figures and Critical Media: Removal of references to Alexei Navalny, independent media outlets (Novaya Gazeta, Echo of Moscow, Radio Svoboda), and opposition activities. Deletions erase documentation of political dissent, protest movements, and critical journalism, creating historical record without meaningful opposition to state policies.

5. Elimination of Sanctions and International Condemnation: Removal of references to international sanctions lists (US, Canada, EU), sanction reasons, and diplomatic isolation. Deletions erase evidence of international community's response to Russian actions, removing context of Russia's pariah status and economic consequences of aggression.

6. Suppression of War Context and Consequences: Removal of terms describing combat operations, occupation, military casualties, and war's impact on civilian life. Deletions eliminate documentation of warfare's reality, replacing detailed conflict description with sanitized administrative language that obscures violence and destruction.

7. Erasure of LGBTQ+ Existence and Rights Discourse: Removal of neutral or positive references to sexual orientation, LGBTQ+ rights, same-sex relationships, and gender identity. While anti-LGBTQ+ content is added, deletions eliminate factual discussion of LGBTQ+ people's existence, international human rights standards, and European Court rulings, creating information void filled only by state propaganda.

Insertion Narratives

1. Legitimization of Territorial Annexation: Systematic insertion of language framing annexed Ukrainian territories (Donetsk, Luhansk, Zaporizhzhia, Kherson) as legitimate subjects of the Russian Federation. Terms emphasize treaties, constitutional incorporation, and administrative structures while noting Ukraine 'disputes' rather than 'owns' these territories. This narrative presents annexation as a completed legal process rather than ongoing occupation.

2. Administrative Normalization of Occupied Territories: Detailed documentation of municipal structures, settlement boundaries, and administrative codes for occupied regions, treating them as routine Russian administrative units. References to Chuvash Republic laws and settlement reorganization create equivalence between legitimate Russian territories and illegally annexed Ukrainian lands, normalizing occupation through bureaucratic language.

3. Justification of 'Special Military Operation': Euphemistic framing of the invasion as a 'special military operation' rather than war. Context suggests military actions are administrative background events rather than aggressive warfare. Terms position Russian actions as procedural operations occurring alongside routine governance activities.

4. Anti-LGBTQ+ Legislative Agenda: Extensive promotion of Russia's expanded anti-LGBTQ+ laws, framing them as protecting children and traditional values. Narrative links LGBTQ+ identity with pedophilia, propaganda, and foreign influence. Emphasizes constitutional priority, legal prohibitions on same-sex marriage, adoption, and gender transition, while citing other restrictive countries as validation.

5. Constitutional Supremacy and Legal Isolationism: Emphasis on Russian constitutional priority over international norms and foreign law. Narrative establishes legal framework where Russian domestic law supersedes external obligations, supporting withdrawal from international human rights frameworks and justifying restrictions on foreign organizations.

6. Foreign Agent and Extremist Designations: Documentation of organizations and individuals labeled as foreign agents or extremist movements. Narrative frames external criticism and domestic opposition as foreign-influenced threats requiring legal suppression. Creates administrative record of state-designated enemies.

7. Comparative Moral Relativism: References to restrictive policies in other countries (DPRK, Saudi Arabia, Central Asia, Africa) to normalize Russian restrictions. Narrative suggests Russia's human rights limitations are internationally common rather than exceptional, deflecting criticism through comparative framing.

Notes: This robustness check uses Claude Sonnet 4.5 with temperature=0.5 to assess sensitivity to temperature parameter.

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 systematic narrative control transforming Wikipedia articles into vehicles for Russian state propaganda. Additions legitimize illegal territorial annexations through bureaucratic language, establish constitutional supremacy over international law, expand anti-LGBTQ+ restrictions, and frame invasion as 'special military operation.' Simultaneously, deletions systematically erase accurate terminology ('invasion,' 'war'), Ukrainian sovereignty markers (administrative codes, local governance), international condemnation (sanctions, court cases), opposition voices (Navalny, independent media), and human rights violations documentation. This bidirectional editing creates alternate reality where: annexations appear legally legitimate; international opposition is minimized; dissent is invisible; LGBTQ+ people are threats; and Russian aggression is reframed as defensive action. The pattern demonstrates coordinated effort to rewrite information about Russia's invasion of Ukraine, political repression, and social policies, transforming encyclopedic content into state-aligned propaganda that obscures violence, eliminates opposition perspectives, and normalizes authoritarian governance and territorial expansion.

Deletion Narratives	Insertion Narratives
1. Suppression of Invasion Terminology: Systematic removal of terms explicitly naming Russian actions as 'invasion' ('вторжение'), 'aggression,' and 'war' against Ukraine. Deletions eliminate language that accurately characterizes the conflict's nature, replacing direct terminology with euphemisms. This suppression obscures Russia's role as aggressor and prevents readers from encountering straightforward descriptions of military aggression. 2. Erasure of Ukrainian Sovereignty Indicators: Removal of Ukrainian administrative codes (KOATUU), postal indices, telephone codes, and references to Ukrainian local councils eliminates evidence of Ukrainian governance structures in occupied territories. These deletions erase administrative infrastructure markers that establish Ukrainian jurisdiction, facilitating narrative that these territories were never functionally Ukrainian. 3. Removal of Territorial Integrity References: Deletion of phrases about Ukraine's territorial integrity, international recognition of borders, and protection of sovereignty eliminates context about violated international norms. These removals obscure that annexations violate Ukraine's internationally recognized borders and global legal frameworks, presenting territorial changes as internal Russian matters rather than international violations. 4. Suppression of Sanctions and International Condemnation: Systematic removal of references to sanctions lists, international legal responses, and diplomatic consequences eliminates evidence of global condemnation. Deletions of terms about European Court, UN resolutions, and country-specific sanctions obscure international community's rejection of Russian actions, creating impression of isolated rather than globally opposed behavior. 5. Elimination of Dissent and Opposition Figures: Removal of references to opposition figures (Alexei Navalny), independent media (Novaya Gazeta, Echo of Moscow, Radio Free Europe), and protest activities erases documentation of internal dissent. These deletions create false impression of domestic consensus, eliminating evidence of Russian citizens and media who opposed government actions or provided alternative information. 6. Erasure of Human Rights Violations: Deletion of terminology about rights violations, freedom restrictions, persecution, imprisonment, and legal proceedings against dissidents removes documentation of authoritarian repression. Removals of references to European Court of Human Rights cases, criminal prosecutions, and civil rights eliminate evidence of systematic suppression of freedoms and judicial persecution. 7. Removal of LGBTQ+ Neutral/Positive Context: While anti-LGBTQ+ content is added, neutral descriptive terms about sexual orientation, same-sex couples, and LGBTQ+ rights are deleted. This asymmetric editing removes factual, non-judgmental information while retaining negative framing, creating exclusively negative representation. Deletions eliminate context that would present LGBTQ+ people as ordinary humans rather than threats.	1. Legitimization of Territorial Annexations: Extensive additions establish annexed Ukrainian territories (Donetsk, Luhansk, Zaporizhzhia, Kherson) as legitimate Russian federal subjects. Terms emphasize legal frameworks, treaties, administrative boundaries, and constitutional authority. This narrative presents annexations as lawful administrative reorganizations rather than contested territorial seizures, creating bureaucratic normalcy around disputed territories. 2. Framing Opposition as Mere Dispute: Terms like 'Ukraine contests' and 'continues to consider' frame international condemnation and Ukrainian sovereignty claims as simple disagreements rather than responses to illegal aggression. This linguistic minimization portrays Russia's position as equally valid, creating false equivalence between aggressor and victim while delegitimizing Ukraine's territorial integrity claims. 3. Administrative Normalization: Detailed technical language about municipal boundaries, settlement types, administrative codes, and local governance structures in occupied territories creates appearance of routine bureaucratic functioning. This administrative veneer masks military occupation, presenting territorial seizure as standard regional reorganization within Russian federal system, obscuring the violent context. 4. Constitutional Supremacy Over International Law: Repeated emphasis on Russian constitutional provisions and domestic legal frameworks establishes primacy of Russian law over international norms. This narrative justifies actions contradicting international law by asserting constitutional authority, particularly regarding annexed territories and social policies, positioning Russia as sovereign arbiter of legality within its claimed borders. 5. Anti-LGBTQ+ Legal Restrictions: Extensive terminology around prohibitions on LGBTQ+ 'propaganda,' same-sex relationships, adoption, and gender transition reflects legislative expansion of restrictions. Framing links LGBTQ+ identities with pedophilia and foreign influence, presenting discriminatory laws as protective measures. This narrative constructs LGBTQ+ visibility as societal threat requiring legal suppression. 6. Foreign Agent and Censorship Frameworks: Terms about foreign agent registries, banned movements, and media restrictions establish legal mechanisms for suppressing dissent and independent information. This narrative portrays restrictions as legitimate security measures against foreign interference, justifying authoritarian controls through framing opposition voices as externally controlled threats rather than legitimate criticism. 7. Special Military Operation Framing: Use of euphemistic 'special military operation' terminology rather than 'war' or 'invasion' reflects official state narrative. Context phrases about military actions present Russian aggression as technical military procedure rather than full-scale invasion. This linguistic framing minimizes severity and moral dimensions of the conflict, normalizing ongoing violence.

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

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 suggest a coordinated shift toward codifying Russia’s territorial claims and legal narratives while suppressing Ukraine-centered, critical, or rights-oriented frames. Additions embed contested regions into RF’s legal-administrative and statistical systems, normalize euphemistic war language, and amplify restrictive domestic policies (anti-LGBT, ‘foreign agents’). Deletions remove ‘invasion/aggression’ terminology, Ukrainian governance markers, sanctions and human-rights references, and oppositional figures/media. Alongside technocratic template changes, the net effect promotes a Russia-centric, legally sanitized portrayal of territory and conflict, narrows space for dissent and sexual content, and aligns encyclopedic structure with current Russian state narratives.

Deletion Narratives	Insertion Narratives
1. Erasing ‘invasion/aggression’ framing of the war: Content using ‘Russian invasion/aggression’ is removed or softened, including references to Donbas war, Russian soldiers, and support for the invasion, reducing culpability language. 2. Removal of Ukraine-centric governance and identifiers: Ukrainian administrative templates and identifiers are purged (KOATUU codes, local councils), diminishing Ukraine’s institutional presence in articles about contested areas. 3. Downplaying sanctions and human-rights accountability: Mentions of sanctions lists, ECHR, human rights violations, and media like Radio Svoboda are stripped, reducing visibility of international/legal pressure on Russia and allies. 4. Reducing visibility of Ukrainian leadership and agency: References to the Ukrainian presidency and Zelensky’s actions are removed, weakening Ukraine’s role as an active subject in the narrative. 5. Censoring sexuality and LGBT-related content: Adult-industry awards, explicit sexual content, and mentions of sexual orientation/gender identity are deleted, aligning with tightening moral/legal restrictions. 6. Pruning Western/independent critics and opposition: Mentions of Alexei Navalny, independent media, and protest narratives are reduced, muting dissenting viewpoints within biographies and event pages. 7. Reformatting away from legacy data and sports minutiae: Postal/telephone codes, climate minutiae, and granular match statistics are removed, replacing older templates with streamlined formats more consistent with current RF-focused standards.	1. Normalizing “new subjects” of the Russian Federation: Edits emphasize Donetsk, Luhansk, Kherson and Zaporizhzhia as new federal subjects, embedding them into RF’s administrative lexicon (oblast, rayon, urban okrug) and population templates, often noting Ukraine’s dispute but asserting Russian composition. 2. Legal-constitutional justification of annexations: Language foregrounds Russia “defining territory,” constitutional priority, and “decisions based on referendum results,” presenting legal veneer for control and integrating these areas into federal law and codes. 3. Russification of administrative geography: Pages gain Russian-style municipal units and boundary changes, including Chuvash Republic statutes, settlement mergers, and urban okrug frameworks—projecting uniform RF administrative practice onto newly claimed and domestic regions. 4. Euphemistic war framing: Edits adopt Russian official terminology (“special military operation,” “military actions in Ukraine”), avoiding ‘invasion’ while acknowledging hostilities and contested status. 5. Expansion of anti-LGBT legal discourse: Increased references to bans on LGBT ‘propaganda,’ adoption by LGBT people, and gender transition, aligning with recent federal restrictions and portraying them as legal norms. 6. Institutionalization of ‘foreign agents’ and banned movements: Growing use of labels like ‘recognized as prohibited in the Russian Federation’ and inclusion in the foreign-agent registry, normalizing stigmatization of organizations and movements. 7. Technocratic standardization and datafication: Articles accrue codes, registries, and dataset references (water-resource codes, GIS releases, sex-disaggregated census), signaling a shift to RF-centric technical templates and statistics.

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

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

Edits suggest a systematic reframing toward Russian legal-administrative narratives: formalizing annexations as new federal subjects, emphasizing accession treaties, and standardizing territorial and municipal structures. Conflict language shifts to the official ‘Special Military Operation’ terminology. Social policy content trends toward conservative legal restrictions, while technical identifiers move to Russian registries. Deletions weaken frames of invasion, occupation, human rights, sanctions, and opposition/media references, and remove explicit sexual content and Ukrainian administrative codes. Overall, the pattern aligns with consolidating Russian state narratives and standards while de-emphasizing international-law, sovereignty, and rights-based critiques.

Deletion Narratives	Insertion Narratives
1. Removal of ‘invasion/war’ framing: Language explicitly calling Russia’s actions an invasion or labeling the conflict a Russia–Ukraine war is removed, reducing accusatory framing and legal responsibility implications. 2. Downplaying sovereignty and occupation claims: Mentions of Ukraine’s territorial integrity, occupation, and the annexation of Crimea are deleted, muting narratives that foreground international-law violations and Ukrainian sovereignty. 3. Pruning opposition/media/sanctions and ECHR references: References to opposition figures and independent outlets, sanctions lists, and appeals to European courts are removed, diluting narratives about political repression, accountability, or international repercussions. 4. Reducing LGBT rights visibility: Rights-based and identity terms related to LGBT communities are deleted, limiting visibility of equality and human-rights frames within relevant articles. 5. Purging pornographic/AVN-related content: Explicit adult-industry material (AVN awards, porn scenes) is removed, aligning with tightening content norms and legal restrictions in Russia. 6. Replacing Ukrainian codes and contact info: Ukrainian administrative identifiers and contact details are deleted, reflecting a shift away from KOATUU, postal, and telephone codes toward Russian systems. 7. Trimming sports minutiae: Match-by-match details, lineups, and minute-level substitutions are removed, suggesting a move to streamline biographies and local pages by cutting event reportage.	1. New RF federal subjects via referendums (Ukraine regions): Additions frame parts of Donetsk, Luhansk, Kherson, and Zaporizhzhia as new subjects of the Russian Federation, often timestamped to September 2022 and justified by referendums. Ukraine’s dispute is noted but positioned as external. The emphasis promotes a legalistic, state-centered justification for territorial changes. 2. Accession treaties as legal basis: Frequent references to ‘treaties’ with Donetsk, Luhansk, Kherson, and Zaporizhzhia present accession as formal, treaty-based acts. This builds a documentary/legal scaffold to legitimize territorial integration under Russian law. 3. Granular inclusion of DNR/LNR localities: Detailed listings of territories and settlements ‘included’ within DNR/LNR and associated oblasts standardize local geography as belonging to Russian-administered units, reinforcing the administrative normality of the changes. 4. Municipal status and boundary reforms (Chuvashia focus): Numerous edits describe laws redefining municipal boundaries, creating/abolishing settlements, and assigning urban/rural status, especially in the Chuvash Republic. This normalizes ongoing internal administrative engineering and harmonizes pages with current Russian regional legislation. 5. Shift to ‘Special Military Operation’ wording: Use of the official term ‘Special Military Operation’ and generic ‘military actions in Ukraine’ reframes the conflict away from ‘war/invasion’ language, aligning with Russian legal and political discourse. 6. Codifying conservative social policy (anti-LGBT, family law): Additions foreground bans on LGBT ‘propaganda’, restrictions around adoption, and family-code references, reflecting tightening conservative policies and their codification in law and enforcement practice. 7. Emphasis on state registries and technical codes: Expanded use of official identifiers and federal agency data (e.g., water-resource codes) signals alignment with Russian technical standards and bureaucratic registries, replacing alternative or legacy systems.

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

The editing pattern suggests a coordinated effort to legitimise Russia’s annexation of Ukrainian territories, reinforce conservative social policies, and emphasise sovereign legal primacy. Simultaneously, language that frames Moscow as an aggressor, highlights human-rights concerns, or foregrounds liberal opposition and LGBT topics is stripped away. Additions craft a narrative of orderly integration and moral righteousness; deletions erase dissenting or critical perspectives. This dual strategy aligns closely with contemporary Russian state discourse aimed at domestic consolidation and international justification of its actions.

Deletion Narratives	Insertion Narratives
1. Characterising the Conflict as a Russian Invasion or Aggression: Terms explicitly calling the 2022 events an “вторжение” or “российско-украинская война” are systematically removed, dulling language that attributes responsibility for hostilities to Russia. 2. Affirming Ukraine’s Territorial Integrity: Phrases about “целостность Украины” or Russia’s “нарушение целостности” disappear, muting references that frame annexations as illegal and stress Kyiv’s sovereign rights. 3. Visibility of Opposition Figures and Independent Media: Mentions of critics like “Алексей Навальный,” outlets such as “Новая газета” and “Эхо Москвы” are deleted, reducing readers’ exposure to domestic dissent and investigative reporting. 4. Coverage of LGBT Identities and Rights: References to “сексуальная ориентация,” “гей/лесбиянка” and LGBT rights vanish, in line with efforts to limit positive or neutral information on queer topics. 5. Details of International Sanctions Against Russia: Information on “санкционные списки,” especially those from the USA or Canada, is pruned, obscuring the extent of punitive measures and their targets. 6. Discussion of Human Rights and Civil Liberties Issues: Removed terms include “права человека,” “право свобода,” and references to the “Европейский суд,” reducing critical context on rights violations or legal recourse. 7. Explicit Sexual Content and Western Adult Entertainment References: Mentions of porn-industry awards (e.g., “AVN”) and explicit acts like “анальный секс” are cut, reflecting a push for conservative moral standards and avoidance of Western erotic culture.	1. Integration of Occupied Ukrainian Territories as New Russian Subjects: Frequent insertions portray Donetsk, Luhansk, Zaporizhzhia and Kherson as “новый субъект” already “входить состав Российской Федерации,” treating them as ordinary federal regions rather than disputed lands. This normalises annexation in readers’ minds. 2. Legal Assertions of Russian Sovereignty Over Disputed Areas: Added phrasing stresses that Russia “продолжает считать территорию” its own and that the “федерация определяет” borders despite the fact that “Украина оспаривает” them, reinforcing Moscow’s legal narrative and downplaying international objections. 3. Municipal & Administrative Re-organisation in Newly Acquired Regions: Large blocks of edits describe the creation, merger or abolition of “сельский поселение,” “городской округ,” and local laws (e.g., Чувашия statutes) to show orderly governance and assimilation of new lands into Russia’s administrative system. 4. Strengthening Anti-LGBT Legislation and Moral Norms: Insertions link LGBT topics with “пропаганда,” “педофилия” and bans on “однополый брак,” referencing the Семейный кодекс and new federal prohibitions. The edits frame anti-LGBT measures as protecting traditional values. 5. Primacy of the Russian Constitution over International Obligations: Phrases such as “приоритет российской Конституции” and “конституция норма” are added to emphasise that domestic law overrides external rulings, supporting recent legislation that rejects foreign courts’ authority. 6. Framing the Conflict as a “Special Military Operation,” Not War: Edits prefer “специальная военная операция” and neutral mentions of “военные действия на Украине,” avoiding the word ‘war’ and aligning with state rhetoric that minimises the scale and legality issues of hostilities. 7. Showcasing Russia’s Global Partnerships Beyond the West: Additions referencing “Саудовская Аравия,” “КНДР,” “страны Африки” and diplomatic roles highlight Moscow’s outreach to the Global South, underscoring a narrative of non-isolation and multipolar alliances.

Notes: This robustness check uses OpenAI O3 with reasoning_effort=medium for cross-model validation.

Table A14: LLM Narrative Assessment (Robustness): Deepseek V3 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

Wikipedia edits reflect a clear pattern of promoting narratives aligned with Russian geopolitical interests, particularly regarding territorial claims, military actions, and internal governance. Additions consistently reinforce state narratives, emphasizing integration of new territories, support for pro-Russian separatist movements, and domestic legal reforms. Conversely, deletions suppress content critical of Russian policies, particularly around military aggression in Ukraine, LGBTQ+ rights, and international condemnation. This editorial pattern aligns with broader state-sponsored narratives aimed at maintaining a favorable portrayal of Russian actions both domestically and internationally.

Deletion Narratives	Insertion Narratives
1. Russian Aggression in Ukraine: Content depicting Russian military actions as invasions or aggressions is frequently removed. 2. Criticism of Russian Policies: Content critical of Russian domestic and international policies is often deleted. 3. Crimean Annexation Controversy: Content that contradicts the Russian narrative of Crimea's annexation is suppressed. 4. LGBTQ+ Rights and Representation: Content supporting LGBTQ+ rights or discussing Russian suppression of LGBTQ+ voices is deleted. 5. Negative Portrayal of Russian Military: Content that negatively portrays Russian military actions or personnel is removed. 6. Ukrainian Sovereignty and Resistance: Content supporting Ukrainian sovereignty and resistance against Russian actions is suppressed. 7. International Condemnation of Russia: Content detailing international condemnation or sanctions against Russia is often deleted.	1. Integration of New Territories: Articles frequently add content about the inclusion of new regions into the Russian Federation, emphasizing territorial claims and governance. 2. Military Actions in Ukraine: Content is added to support narratives around Russian military actions, focusing on strategic gains and operations. 3. Cultural and Administrative Changes: Additions highlight changes in local governance, cultural policies, and administrative reforms within Russia. 4. Pro-Russian Separatist Movements: Content supports narratives around pro-Russian separatist movements in Ukraine, emphasizing their legitimacy and governance. 5. International Relations and Sanctions: Additions focus on Russia's stance against international sanctions and its foreign policy maneuvers. 6. Internal Governance and Law: Content emphasizes domestic legal reforms and governance structures, reinforcing state control. 7. Historical and Territorial Claims: Additions reinforce historical narratives and territorial claims, aligning with Russian geopolitical interests.

Notes: This robustness check uses Deepseek V3 (via Openrouter) with temperature=1 for cross-model validation.

G Narrative Placebo: Wikipedia changes over time

This appendix section reports the placebo narrative-extraction results for Wikipedia changes over time between the fork and the 1st of January 2025 (Table A15). We also verify that the changes in Ruwiki narratives over time between the fork and the 1st of January 2025 are very close to the comparison between Ruwiki and Wikipedia on the 1st of January 2025 (see Table A16 presented below as compared to 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 in the three panels below is reproduced verbatim from the model output and has not been edited by the authors.

Overall Assessment

The additions appear to focus on expanding narratives around historical, cultural, and sports achievements, as well as documenting rural and infrastructural developments. This suggests a trend towards bolstering national pride and significance. The deletions, on the other hand, indicate a potential suppression of detailed official data and contested region information, possibly to control both domestic and international discourse. There is also a noticeable trend of removing detailed demographic and geographic descriptors, which aligns with a reduction in information about sensitive topics or politically charged areas.

Deletion Narratives	Insertion Narratives
1. Suppression of Official Codes and Data: Removals related to official geographic codes and federal data suggest a potential effort to limit public access to detailed governmental and geographic data. 2. Reduction in Specific Census Data: Decreased documentation of detailed census and demographic data could indicate a decision to streamline or limit intricate population data. 3. Minimization of Religious Object Information: Content about religious objects and structures is being removed, possibly reflecting sensitivities or changes in how religious information is shared. 4. Limitation on Data from Contested Regions: Information pertaining to disputed areas such as Nagorno-Karabakh is being reduced, likely due to the political sensitivities surrounding these regions. 5. Cultural and Historical Information Suppression: Reductions in literature and historical figure information suggest a shift in focus or a censuring of certain cultural narratives. 6. Removal of Sanction-Related Information: Details about sanctions and related geopolitical actions are being removed, potentially to control international perceptions and minimize focus on these topics. 7. Simplification of Demographic and Geographic Descriptions: Simplifying or removing precise demographic and geographic data reflects a possible strategic move to generalize information made available to the public.	1. Expansion of Historical Conflicts: Content about Russian involvement in historical wars and conflicts, such as the World Wars and regional invasions, seems to be increasingly added. This reflects an effort to document Russia's military history in greater detail. 2. Development and Infrastructure: There is significant emphasis on rural development and infrastructure, indicating an interest in documenting economic advancements in rural areas and the expansion of infrastructure. 3. Sports Achievements: Additions related to sports, championships, and Russian athletes reflect a narrative focused on promoting national pride through sports achievements. 4. Cultural Heritage and Legacy: Content related to cultural heritage, historical sites, and significant cultural figures are being expanded, suggesting a focus on preserving and promoting Russian cultural narratives. 5. Rural and Agricultural Development: There is increased documentation of rural and agricultural settings, suggesting a focus on the importance of agriculture in the Russian economy and lifestyle. 6. Political and Legal Processes: Additions around laws, political decisions, and legal frameworks suggest a focus on recording Russia's political processes and regulation changes. 7. Religious and Social Structures: There is a notable increase in content about religious and social structures, indicating an interest in documenting the diverse religious practices and community compositions in Russia.

Notes: This table reproduces verbatim the output of the GPT-4o narrative-extraction pipeline with temperature set to 1. 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 within-Wikipedia placebo examines changes within Wikipedia over time (2023-05-01 to 2025-01-01), tracking how content evolved during this period. The LLM identified 7 distinct narratives for deletions and 7 for insertions based on statistically significant term associations (bigrams and trigrams only). The overall assessment synthesizes patterns across all narratives. Narratives are generated from the top 500 terms by χ^2 statistic in each direction.

Table A16: LLM Narrative Assessment: Ruwiki over time, January 2025 vs. the fork

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

Overall Assessment

Analyzing the patterns in Wikipedia edits reveals an effort to reinforce narratives around Russian territorial integration and administrative control, promoting the legitimacy of new territories within the Russian Federation. Conversely, there is a systematic removal of content related to Ukrainian sovereignty, LGBTQ+ rights, and opposition figures, suggesting a push to suppress these subjects. The editing patterns reflect a narrative strategy attempting to align public information with certain political and cultural goals, emphasizing state narratives while minimizing dissenting perspectives, especially regarding contested sovereignty and civil liberties.

Deletion Narratives	Insertion Narratives
1. Suppression of Ukrainian Governance: Deletions focus on removing references to Ukrainian administrative structures in contested regions. 2. Obfuscation of Conflict Details: Content related to the Russian invasion and Ukrainian territorial integrity is being deleted, possibly to reduce visibility. 3. Erasure of Ukrainian Sovereignty: References to Ukrainian territorial claims and sovereignty are being removed from articles. 4. Removal of Opposition Figures: Deletions include content related to opposition figures and independent media, possibly to limit dissenting perspectives. 5. LGBTQ+ Rights Suppression: Content related to LGBTQ+ rights, activism, and sexual orientation is being removed, potentially to suppress these topics. 6. Censorship of Free Expression: Terms linked to freedom of speech and protests are being deleted, possibly to control narratives around civil liberties. 7. Concealment of Conflict Impact: Deletions aim to reduce references to civilian impacts, rights abuses, and international responses to the conflict.	1. Integration of New Territories: Content additions focus on territorial changes involving Donetsk and Luhansk, framing these areas as integrated subjects of the Russian Federation, often contested by Ukraine. 2. Legitimization of Republics: Additional content includes narratives that legitimize the Donetsk and Luhansk People's Republics, mentioning their inclusion in official Russian frameworks. 3. Legal Assertions Over Contested Areas: Content focuses on defining the legal status of territories as part of Russia, in contrast to Ukrainian claims. 4. Formal Agreements for New Boundaries: Addition of terms suggests narratives around formal agreements and treaties concerning the new territorial boundaries. 5. Military Operations Narrative: Content additions relate to the Russian military actions in Ukraine, framing them within a specific operational context. 6. Cultural and Administrative Integration: New content focuses on cultural assimilation and administrative changes within newly integrated regions of Russia. 7. Territorial Status Reinforcement: Additions reinforce the narrative of disputed areas' statuses as established elements of the Russian Federation.

Notes: This table reproduces verbatim the output of the GPT-4o narrative-extraction pipeline with temperature set to 1. 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 within-Ruwiki analysis examines changes within Ruwiki over time (2023-05-01 to 2025-01-01), tracking how content evolved during this period. The LLM identified 7 distinct narratives for deletions and 7 for insertions based on statistically significant term associations (bigrams and trigrams only). The overall assessment synthesizes patterns across all narratives. Narratives are generated from the top 500 terms by χ^2 statistic in each direction.

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.³³ 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.

For each bootstrap iteration $b = 1, \dots, 100$, 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 β by minimizing the penalized negative log-likelihood:

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

where $\pi_i = \Pr(y_i = 1 \mid \mathbf{x}_i) = \frac{\exp(\beta_0 + \mathbf{x}_i' \beta)}{1 + \exp(\beta_0 + \mathbf{x}_i' \beta)}$ represents the predicted deletion probability for article i , and $\lambda \geq 0$ controls the strength of the L1 penalty applied to the coefficient vector.³⁴

³³N-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.

³⁴We estimate the models using the `glm1r` package in R, which implements efficient coordinate-descent algorithms for high-

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 that appear as nonzero predictors in at least 5% of the bootstrap samples, focusing on terms whose predictive power persists across alternative realizations of the control group. After completing the statistical analysis, we translate the identified Russian terms into English to facilitate interpretation and presentation of the findings.

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.

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 $\lambda_{\min}$.

Table A17: LASSO Deletion Prediction Results: Top 15 Predictors by Specification

Contributor-Risk Notice Terms Included				Contributor-Risk Notice Terms Filtered			
#	Term	Coef.	Broad topic	#	Term	Coef.	Broad topic
<i>Article Categories</i>							
1	Pornography articles	6.50	Culture	1	Pornography articles	6.50	Culture
2	year in Ukraine	4.55	War	2	year in Ukraine	4.55	War
3	articles with editnotice	3.40	Other	3	articles with editnotice	3.40	Other
<i>Category Terms</i>							
1	pornography article	6.72	Culture	1	pornography article	6.69	Culture
2	LGBT movement	4.55	Culture	2	LGBT movement	4.51	Culture
3	Russian invasion	3.97	War	3	Russian invasion	4.05	War
4	invasion Ukraine	3.18	War	4	Russian war	3.14	War
5	Russian war	2.82	War	5	invasion Ukraine	3.11	War
6	caution editing	2.81	War	6	battle territory	2.54	War
7	war crime	1.66	War	7	war crime	1.83	War
8	battle territory	1.39	War	8	Russian Federation	0.42	Politics
9	English language	0.33	Other	9	English language	0.30	Other
<i>Title Terms</i>							
1	Russian invasion	5.25	War	1	Russian invasion	5.25	War
2	pornography article	4.62	Culture	2	pornography article	4.62	Culture
<i>Category Terms (with unigrams)</i>							
1	pornography	6.22	Culture	1	pornography	5.73	Culture
2	sexual	3.95	Culture	2	sexual	3.95	Culture
3	LGBT	3.22	Culture	3	LGBT	3.22	Culture
4	activist	1.99	Politics	4	activist	1.99	Politics
5	invasion	1.94	War	5	invasion	1.92	War
6	Russian	1.55	Politics	6	Russian	1.64	Politics
7	caution	1.37	War	7	Donetsk	1.39	War
8	Donetsk	1.31	War	8	public	1.31	Politics
9	public	1.30	Politics	9	person	1.27	Politics
10	person	1.25	Politics	10	Ukraine	1.15	War
11	Ukraine	1.15	War	11	woman	0.89	Culture
12	woman	0.88	Culture	12	article	0.78	Other
13	mass	0.73	Politics	13	mass	0.70	Politics
14	Russia	0.56	Politics	14	Russia	0.57	Politics
15	pornography article	0.52	Culture	15	political	0.52	Politics
<i>Title Terms (with unigrams)</i>							
1	pornography	4.17	Culture	1	pornography	4.17	Culture
2	battle	3.59	War	2	battle	3.59	War
3	Ukraine	2.72	War	3	Ukraine	2.72	War
4	Russian (adjective)	2.56	Politics	4	Russian (adjective)	2.56	Politics
5	Russia	1.56	Politics	5	Russia	1.56	Politics

Notes: This table presents the top 15 deletion predictors from bootstrapped LASSO logistic regression for each specification. Coefficients represent log-odds ratios systematically shrunk toward zero by L1 regularization. Terms appearing in fewer than 75% of bootstrap iterations are excluded. Default mode includes contributor-risk notice terms; No-contributor-risk-notice mode filters these terms during n-gram generation. Categories: Culture (pornography, LGBTQAI+ topics), War (Ukraine invasion, conflict), Politics (Russian political figures, activists), Other (miscellaneous).

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

Contributor-Risk Notice Terms Included				Contributor-Risk Notice Terms Filtered			
#	Term	Coef.	Broad topic	#	Term	Coef.	Broad topic
<i>Title+Category Terms (with unigrams)</i>							
1	pornography	5.80	Culture	1	pornography	5.41	Culture
2	sexual	3.35	Culture	2	sexual	3.34	Culture
3	LGBT	2.35	Culture	3	LGBT	2.35	Culture
4	activist	2.19	Politics	4	activist	2.20	Politics
5	invasion	1.28	War	5	invasion	1.27	War
6	caution	1.13	War	6	Russian	1.13	Politics
7	Russian	1.06	Politics	7	Donetsk	1.07	War
8	public	1.05	Politics	8	public	1.06	Politics
9	Donetsk	1.04	War	9	person	1.01	Politics
10	person	1.00	Politics	10	Ukraine	0.99	War
11	Ukraine	0.99	War	11	Russian war	0.82	War
12	Russian war	0.81	War	12	article	0.78	Other
13	mass	0.67	Politics	13	mass	0.65	Politics
14	war crime	0.65	War	14	war crime	0.64	War
15	woman	0.62	Culture	15	woman	0.62	Culture

Notes: This table presents the top 15 deletion predictors from bootstrapped LASSO logistic regression for each specification. Coefficients represent log-odds ratios systematically shrunk toward zero by L1 regularization. Terms appearing in fewer than 75% of bootstrap iterations are excluded. Default mode includes contributor-risk notice terms; No-contributor-risk-notice mode filters these terms during n-gram generation. Categories: Culture (pornography, LGBTQAI+ 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).

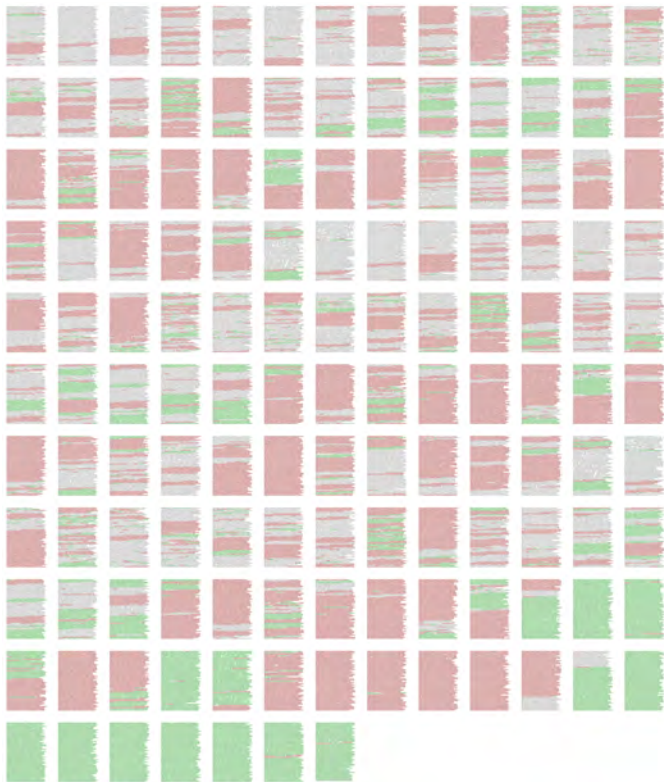


Figure A10: Example of heavy edits: *Human rights in Russia* article (domestic-politics pillar).



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 fixed effects and β is the coefficient vector of interest. The estimates solve the penalized maximum-likelihood problem

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

The article fixed effects absorb all article-level variation in editing intensity and topic, so that $\hat{\beta}_j$ captures the systematic association between an n-gram and the *direction* of editorial change *within* an article. The penalty parameter λ is selected by 10-fold cross-validation on out-of-sample deviance.

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. This gives rise to three conceptually distinct groups of terms.

Taxonomy of terms selected by the editing LASSO.

- (i) **Terms selected in both comparisons (in the opposite direction): staleness or censorship by omission.** Terms that characterize both the R_T vs. W_T and the W_T vs. W_F differences 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 not incorporating Wikipedia’s updates. In the frozen-clone benchmark, where Wikipedia moves with the world while Ruwiki remains at its fork-date state, the two comparisons are identical (with the opposite sign) and

would select the same terms for deletions and insertions (but deletions would become insertions and vice versa). Even though staleness terms may also include censorship, when only new information sensitive for the regime is not updated (as mentioned above), empirically it is impossible to separate active censorship of new events that occur between F and T from passive staleness.³⁵

- (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 do not reflect 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 that are 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.

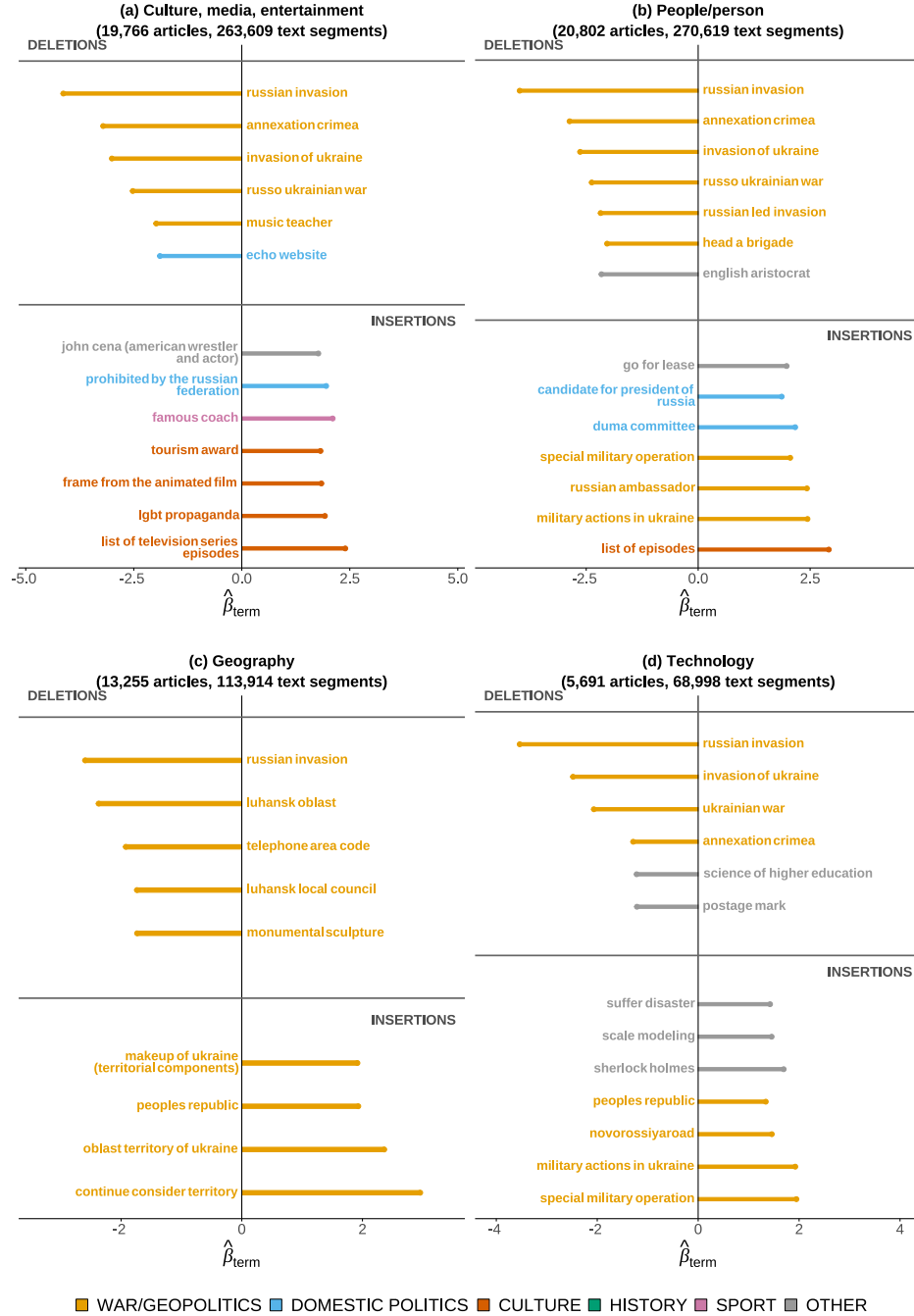
In Appendix Section L.2 (below), we assess the discriminating power of the selected vocabulary using held-out classification. The vocabulary distinguishes Ruwiki from contemporaneous Wikipedia substantially better than it distinguishes contemporaneous Wikipedia from its fork-date version.

³⁵In principle, the same terms could be selected in both comparisons with the same sign. For example, this could occur if Wikipedia introduces a concept in response to a particular event, while Ruwiki at T discusses that event even more extensively, indicating varying-intensity updating of the same events across platforms. In practice, this category is empirically negligible.

³⁶This 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

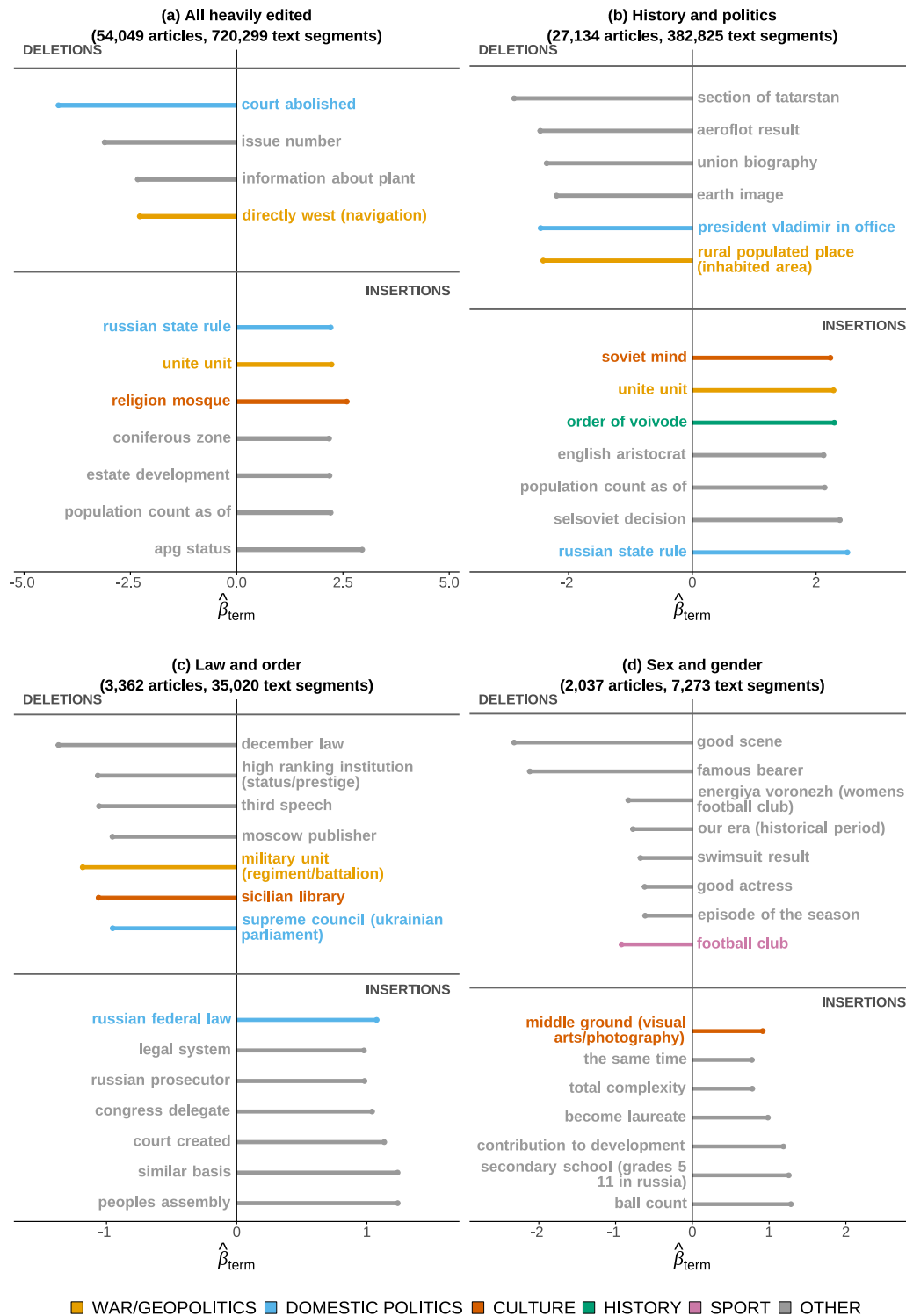


Note: Baseline editing LASSO. Additional subsamples by LLM category. Estimated logit LASSO coefficients ($\hat{\beta}_{term}$) from a regularized logistic regression predicting text additions (positive) versus deletions (negative) by Ruwiki editors. Models are estimated separately by topic using lemmatized bigrams/trigrams and article fixed effects. All models use 10-fold cross-validation. Bar colors indicate semantic manually coded umbrella categories. WAR/GEOPOLITICS (gold/yellow), DOMESTIC POLITICS (blue), CULTURE (orange), HISTORY (green), SPORT (purple), OTHER (gray). (No terms in this figure are classified as HISTORY or SPORT.) Sport is added as an additional umbrella category beyond the five used for deleted articles. Terms are ordered hierarchically by the maximum absolute coefficient within umbrella categories. The unit of observation is a text segment.

L.1 Within-Wikipedia Temporal Placebo: Editing LASSO

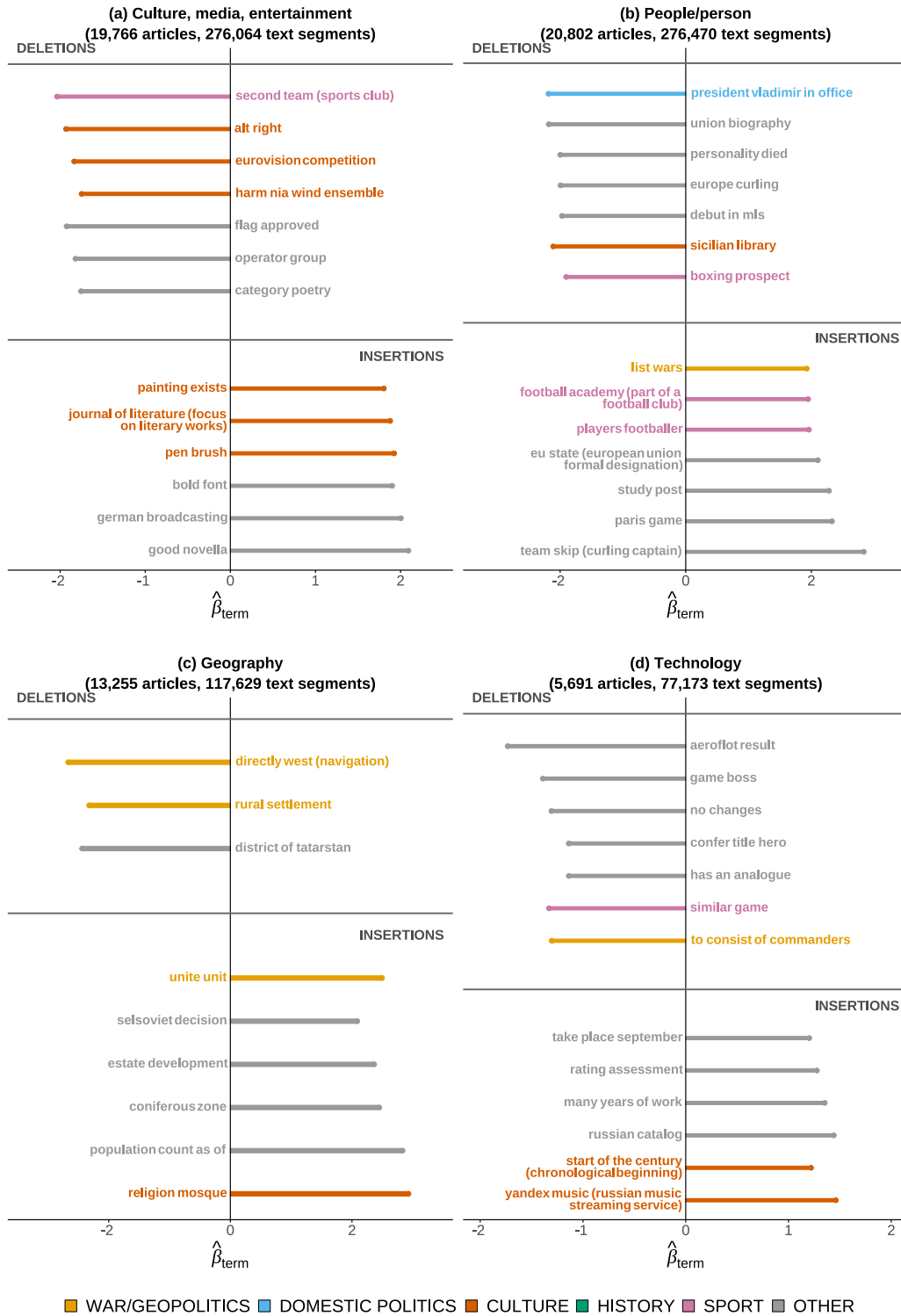
This section presents the within-Wikipedia temporal placebo for the editing LASSO, paralleling the cross-sectional analysis in Section VI.B. The temporal placebo compares Russian-language Wikipedia in May 2023 against Russian-language Wikipedia in January 2025. Figure A14 presents terms chosen by LASSO in within-Wikipedia placebo and not in Ruwiki-Wikipedia baseline. Table A19 presents terms chosen by LASSO in both samples.

Figure A14: Wikipedia Placebo: LASSO predictions of deletions and insertions in heavily edited articles, main LLM topic categories



Note: Estimated logit LASSO coefficients ($\hat{\beta}_{\text{term}}$) from a regularized logistic regression predicting text additions (positive) versus deletions (negative) by Wikipedia editors between the fork and Jan 2025. Models are estimated separately by topic using lemmatized bigrams/trigrams and article fixed effects. All models use 10-fold cross-validation. Bar colors indicate semantic manually coded umbrella categories: WAR/GEOPOLITICS (gold/yellow), DOMESTIC POLITICS (blue), CULTURE (orange), HISTORY (green), SPORT (purple), OTHER (gray). Sport is added as an additional umbrella category beyond the five used for deleted articles.

Figure A15: Wikipedia Placebo: LASSO predictions of deletions and insertions in heavily edited articles, additional LLM topic categories



Note: Estimated logit LASSO coefficients ($\hat{\beta}_{\text{term}}$) from a regularized logistic regression predicting text additions (positive) versus deletions (negative) by Wikipedia editors between the fork and Jan 2025. Models are estimated separately by topic using lemmatized bigrams/trigrams and article fixed effects. All models use 10-fold cross-validation. Bar colors indicate semantic manually coded umbrella categories: WAR/GEOPOLITICS (gold/yellow), DOMESTIC POLITICS (blue), CULTURE (orange), HISTORY (green), SPORT (purple), OTHER (gray). Sport is added as an additional umbrella category beyond the five used for deleted articles.

Table A19: Terms picked by both the baseline Ruwiki–Wikipedia comparison and the placebo Wikipedia over-time comparison editing LASSO, by direction of overlap

	Common terms	
	Opposite direction (staleness or censorship by omission)	Same direction (varying-intensity updating)
LLM Category:		
All heavily edited	official code; religious facility (building used for worship); religion mosque; rus rule; section of tatarstan	koatuu code (ukrainian administrative code)
History and politics	religious facility (building used for worship); rus rule; to cede territory; section of tatarstan	koatuu code (ukrainian administrative code)
Law and order	annexation crimea; house of soviets	
Sex and gender	invasion of ukraine; famous bearer	
Culture, media, entertainment	alt right; dance dancer; breakdancer; soviet mind	
People	english aristocrat; second team (sports club); multiple prizewinner; football academy (part of a football club)	
Geography	study code; official code; district of tatarstan (official administrative unit); religious facility (building used for worship); religion mosque; to cede territory; section of tatarstan; linguistic population	koatuu code (ukrainian administrative code)
Technology	game boss; plot description; game feature; open plant	

Note: These terms were selected by the editing LASSO in both the baseline Ruwiki–Wikipedia comparison and the within-Wikipedia placebo comparison. As explained in Appendix Section E, selection in opposite directions is consistent with staleness or censorship by omission, whereas selection in the same direction reflects updating of different intensity in response to the same real-world event (see footnote 35). Here, the relevant event was the replacement of Ukraine’s KOATUU administrative classification code with the KATOTTH code.

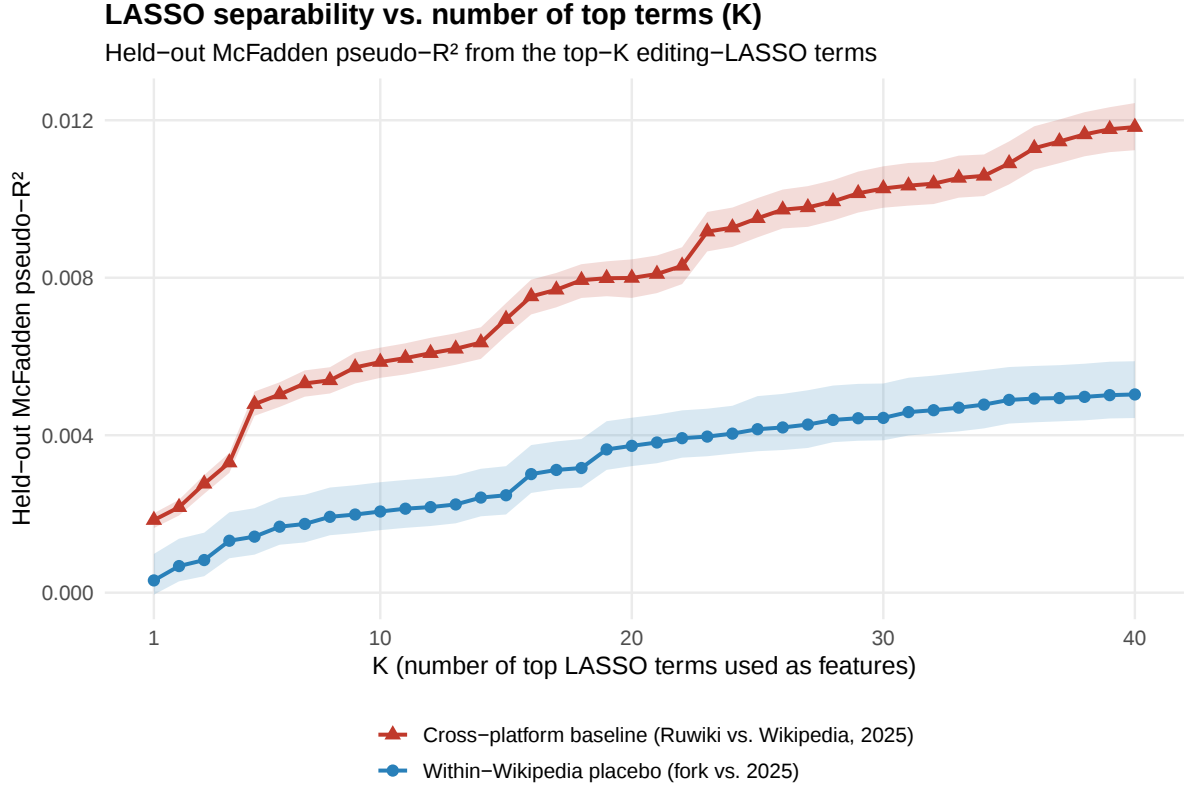
L.2 Competitive LASSO selection and separability

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 A16 reports the results for $K = 1, \dots, 40$. 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 A16: 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 ($K = 1, \dots, 40$). 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 A20: Terms most characteristic of the Ruwiki–Wikipedia baseline comparison only (signed Cramér's V)

All heavily edited	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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 A21: Terms most characteristic of the within-Wikipedia placebo comparison only (signed Cramér's V)

All heavily edited	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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	
<i>Deletions</i> 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	<i>Insertions</i> 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 A22 and presents the corresponding results for edited but not heavily edited articles in Table A23.

Contributor-risk notices are rare but present in both 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: Summary statistics for heavily edited articles

	Wikipedia		Ruwiki		Mean difference	Articles
Outcome	Mean	SD	Mean	SD	R-W	
<i>Panel A: Without contributor-risk notice</i>						
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
<i>Panel B: With contributor-risk notice</i>						
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 A23: Text length and readability in edited but not heavily edited articles

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

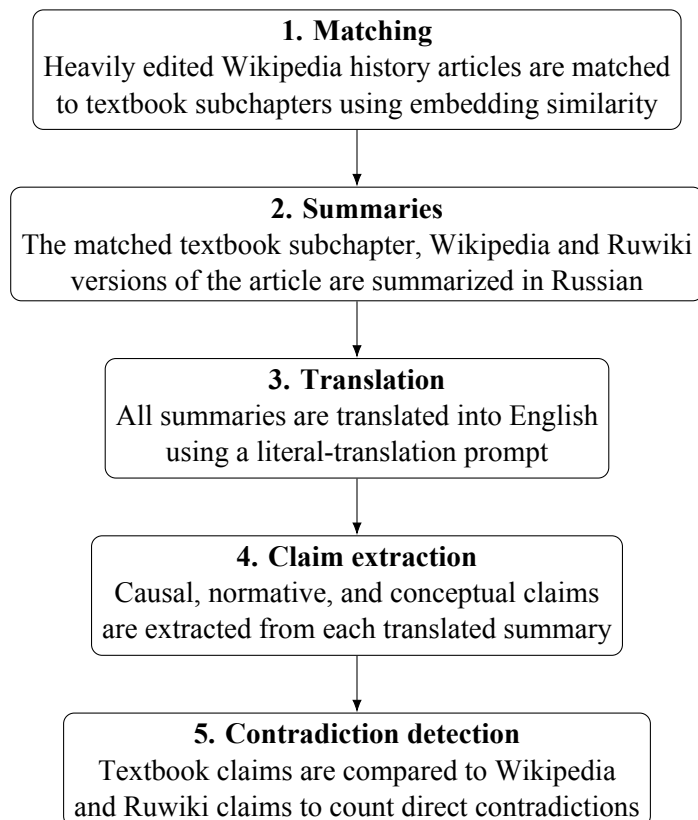
Notes: The sample consists of edited but not heavily edited articles observed in both Wikipedia and Ruwiki. A contributor-risk notice warns Wikipedia editors that contributing to the article may expose them to legal or personal risk in Russia or Belarus and advises them to conceal their identity. Panels 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 A17.

Figure A17: 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 log article length and the number of claims extracted from articles and textbook subchapters. 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
<i>Panel A. Exact paired tests</i>						
Any contradiction	0.060	0.190	0.130	15	2	0.002
Contradiction count	0.080	0.250	0.170	15	3	0.008
<i>Panel B. Within-pair permutation tests</i>						
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.

O.4 Examples of Contradictions

Table [A26](#) reports examples of contradictions between Wikipedia articles and the official history textbooks.

Table A26: Largest Wiki–Ruwiki Differences in Textbook Contradiction Classifications

<i>Cos</i>	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.