Tracking Wikipedia Weaponization and Culture Heritage Manipulation in the Russo-Ukrainian context

Maxime Garambois

Supervisors: Dr. Hamest Tamrazyan, Dr. Emanuela Boros
Digital Humanities Laboratory
Director: Prof. Dr. Frédéric Kaplan

January 9, 2026

Contents

1 Introduction

1.1 Motivation

Since its creation in 2001, Wikipedia has become one of the most epistemic reliable digital platform in the world [1], attracting almost 2 billions of monthly readers in 2025 seeking information on current and historical events [2]. Its open "wiki" model allowing anyone to edit, constitutes both its greatest strength and its main vulnerability [3]. On the one hand, a vast community of daily editors contributes to the expansion of global human knowledge. On the other hand, Wikipedia is also exposed to large-scale campaigns of misinformation [4], manipulation [5] and more recently, the weaponization of cultural heritage. This study investigates how Wikipedia can function as a platform for "weaponization" edits, understood as the strategic manipulation, distortion, or erasure of cultural identity, memory, and heritage for political or military purposes. Since early 2025, the idea of tracking the weaponization and manipulation of cultural heritage in Ukraine through student semester projects has been developed and supervised by Hamest Tamrazyan and Emanuela Boros. During the first part of the year, an initial semester project was carried out by Mohamed Hidi Hedri. The objective of this project was to classify edits from a corpus of 320 Ukraine-related cultural Wikipedia articles as either *weaponizing* or *not weaponizing* using a large language model (LLM). Based on this dataset, the project identified the most frequently targeted thematic areas and proposed a taxonomy of weaponization types to refine subsequent analyses.[6].

1.2 Goal

The purpose of this project is to have a proper dataset that can be used in the future for training models that detect weaponized content in the context of the 2026 CROSS project[1]. This requires comprehensive preprocessing and deep examination of the dataset. This report outlines the methods used to preprocess and analyse the data 2.1, with the aim of enhancing our understanding of its structure and revealing meaningful patterns. Combined with Hidi's project, we aimed to answer four questions :

- **What?** Which types of content or topics related to cultural heritage are primarily targeted?
- **How?** Which techniques—semantic, structural, or algorithmic—are employed by actors who engage in the weaponization of content?
- **Who?** Who are the users responsible for weaponizing content, and what are their main characteristics?
- **Why?** What motivations drive users to manipulate content related to cultural heritage?

The previous phase of the research project has already partially addressed the "*what*" and "*how*" questions. Consequently, this project initially focused on addressing the "*who*" question; however, as the investigation progressed, it also revealed new insights into the types of content being targeted and the techniques used to manipulate it.

1.3 Challenges

Joining an ongoing project presents several methodological challenges. It first requires developing a clear understanding of the work already completed, while simultaneously handling and restructuring large, unfamiliar datasets and identifying the expected analytical outcomes necessary to move the project forward. Throughout this study, working with Wikipedia edits and metadata proved to be particularly demanding.

[1] `https://www.epfl.ch/schools/cdh/cross-2026/`

Although the Wikimedia Action API[2] provides extensive access to information about users, edits, and articles, its documentation is often incomplete or insufficiently detailed. As a result, additional effort and time were required to identify the appropriate endpoints and parameters needed for specific analyses.

A major and recurrent challenge across nearly all feature analyses concerns the limited availability of temporal metadata. For many features of interest, such as user group membership, article quality assessments, or protection status, information is generally accessible only in its current state, rather than as it existed at the time an edit was made. Building a reliable dataset therefore requires reconstructing historical contexts that are often missing, inconsistently recorded, or buried within extensive Wikipedia logs.

Finally, working with raw Wikipedia text and data from raw collected text introduces substantial preprocessing complexity. Article content and edit comments require extensive cleaning and normalisation using custom processing.

2 Data collection and structure

2.1 Data collection

A former data collection has already been created in Hidi's project, where they built a so called fine grained database with edits already labeled with a specific taxonomy. However, the authors of those edits were unknown and lost in the fine grained step. In order to fill this gap, a match making processus has been conducted to retrieve the user name.

2.2 Structure

This semester project report is divided into 3 complementary parts:

1. **Weaponized vs. Non-Weaponized Analysis :** A global overview of editing patterns among a small and a large database of 2,336 and 19,769 unique users respectively, distinguishing between edits classified as *Weaponized* or *Not Weaponized*.
2. **Weaponizing Users — Fine-Grained Analysis :** A focused investigation of 400 unique users from the fine-grained dataset, restricted only to edits identified as *Weaponized* by the language model.
3. **Article Analysis :** A detailed examination of Wikipedia governance, community processes, and policy structures relevant to user behavior and content moderation.

Alongside this report, a Jupyter Notebook has been developed to provide additional plots, detailed statistical analyses, and extended visualisations.

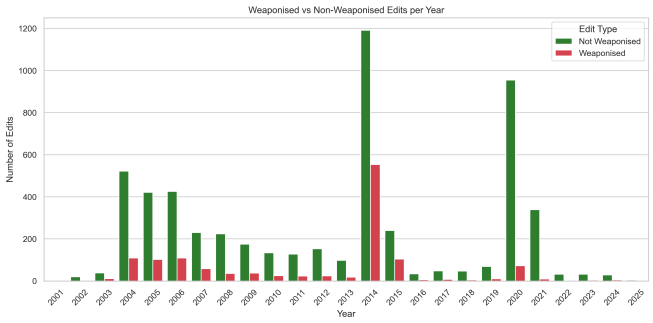
3 Weaponized vs. Non-weaponized Analysis

This section analyzes user- and edit-level patterns to identify robust, reusable features for training future models to detect weaponized edits. This section and the following ones are systematically designed to identify patterns of weaponization based on the user type and additional features. We distinguish three categories of users: registered users, who hold a Wikipedia account; unregistered users, who are identifiable only through their IP addresses; and bots, which are non-human actors and are therefore less analyzed in this report. The comparative analysis between registered and unregistered users, often in combination with other features, sits at the core of the analytical framework of this work.

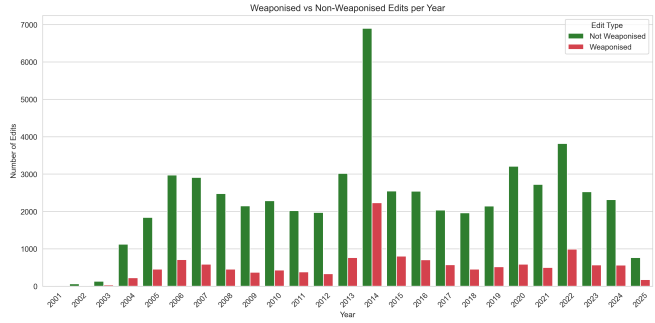
[2] `https://www.mediawiki.org/wiki/API:Action_API`

It is important to note that edits classified as weaponizing or non-weaponizing by the LLM do not constitute ground truth, but rather heuristic labels. Only expert manual annotation can provide definitive validation.

3.1 Datasets Overview



(a) weaponized vs non-weaponized edits over time [small db].



(b) weaponized vs non-weaponized edits over time [large db].

Figure 1: Relationship between edit frequency and major geopolitical events involving Russia and Ukraine across both databases. Notable periods include the 2004–2005 Orange Revolution and subsequent change of presidency, the 2013–2015 Russian annexation of Crimea, the 2019–2020 COVID-19 pandemic, and the ongoing Russian invasion of Ukraine beginning in 2022.

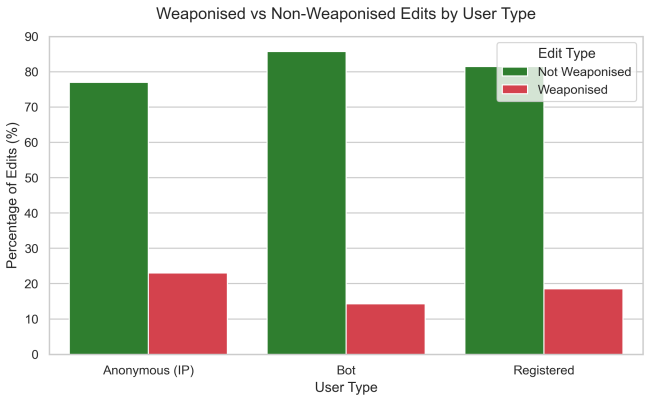
Metric	Small	Large	Change	% Change
Total edits	6922	69908	+62986	+910.1%
Total unique users	2336	19769	+17433	+746.4%
Number of unique articles	40	40	0	0%
Bot users	91	236	+145	+159.3%
Anonymous users	966	10609	+9643	+998.2%
Registered users	1279	8924	+7645	+597.7%
Bot edits	287	3910	+3623	+1262.7%
Anonymous edits	1361	17820	+16459	+1208.9%
Registered edits	5274	48178	+42904	+813.6%
weaponized edits	1333	13445	+12112	+908.5%
Non-weaponized edits	5589	56463	+50874	+910.2%

Table 1: Comparison of summary statistics of small and large datasets.

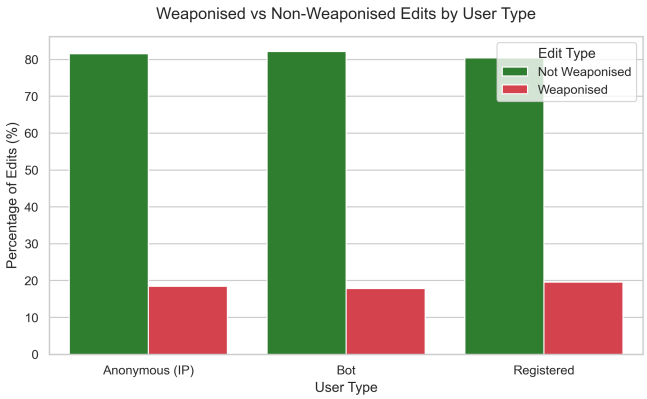
The small database represents a subset of the larger dataset, and the comparison between the two is intended to assess whether the patterns observed in the smaller sample are consistent with and representative of those in the full dataset. As shown in Table 1, the proportion of weaponizing and non-weaponizing edits remains stable across both datasets, with approximately 19.2% of all edits classified as weaponising, attesting that the small database is a reliable sample for weaponization detection. The Figure 1 displays for the large database a more harmonise and coherant distribution of edits according to the major geopolitical events as well as a more coherant and interessant edits/article distribution (see Annex D).

3.2 Data Exploration

First thing that is straightforward to check is the amount of weaponizing vs non weaponizing edits per user type.



(a) weaponized vs non-weaponized edits distribution for registered, unregistered and bot users. [small db]



(b) weaponized vs non-weaponized edits distribution for registered, unregistered and bot users. [large db]

Figure 2: Comparison of weaponizing vs non weaponsing edits distribution across the different user types.

To assess whether the distribution of weaponized versus non-weaponized edits differs across user types, we performed a Chi-square test of independence using the `scipy.stats` library. The contingency table includes three user categories (Registered, Anonymous (IP), Bot) and two edit types (Weaponized, Not Weaponized).

More explorations were also conducted, in particular a visualization through heatmaps of the editing activity over the years and over the week days for each user type. We wanted to visually understand when users are editing, due to which event, and if distinction can be done regarding the user type. We also wanted to know whether the comments left to justify and support an edit have an impact for the weaponization detection. Results are available in the book.

3.3 Users contributions, metadata and motivations analysis

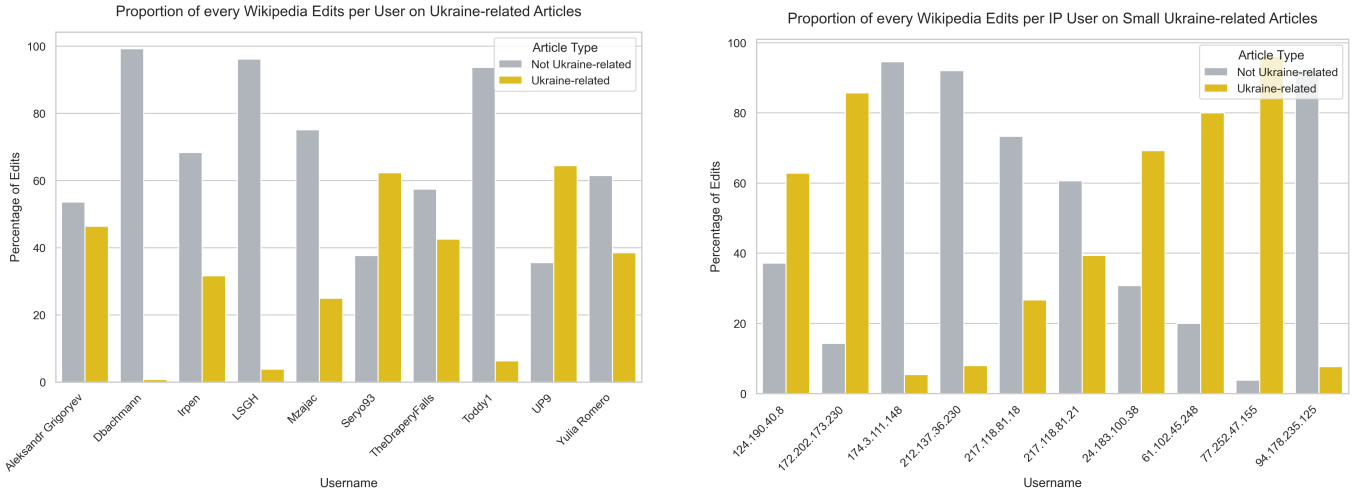
3.3.1 Users contributions

In order to retrieve user **contributions**—defined as all revisions made by a user since it's account registration date—we used the Wikimedia API via the `action=query` endpoint with `list=usercontribs` and `ucprop=ids|title|timestamp|comment|sizediff|flags|tags|oresscores`, allowing us to collect complete edit histories.

We only focused on the ten most active registered users, who together account for approximately 31% of all registered-user edits (1,634 out of 5,274 for small database) and the ten most active Anonymous (IP) users. The total number of edits made by the top ten IP users is relatively small, which limits the depth of the analysis. However, since the computational pipeline was already in place for registered users, we can still visualize these data for completeness.

Using the `ucprop=title`, we analyzed whether their editing activity was specifically concentrated on Ukraine–Russia–related topics or other subjects. For each of these users and in order to determine if an article title is related to Ukraine or Russia, all edited article titles were extracted and classified as Ukraine- or Russia-related based on the presence of predefined keywords. Two keyword lists were constructed: a *narrow* list containing core Ukraine- and Russia-specific terms, and a broader list including additional political, geographical, military, and historical terms. Both lists were initially generated using ChatGPT (GPT-5.0;

prompt provided in the Annex A and subsequently validated by a domain expert. Figure 3 presents the results obtained using the narrow keyword set, while the analysis based on the broader set is reported in Annex A.



(a) Proportion of Ukraine–Russia–related edits for the top ten registered users, based on the small keyword set.

(b) Proportion of Ukraine–Russia–related edits for the top ten IP users, based on the small keyword set.

Figure 3: Comparison of the proportion of Ukraine–Russia–related edits (in yellow) among the top ten registered and IP users. [small db]

Registered users. We observe that some users, such as **Dbachmann**, **LSGH**, and **Toddy1**, edit articles that are not directly related to Ukraine or Russia more than 90% of the time. Other users, including **Irpen**, **Yulia Romero**, and **Mzajac**, show moderate engagement with Ukraine–Russia topics, with approximately 30% of their edited articles falling into this category. **Aleksandr Grigoryev** and **TheDraperyFalls** display a more balanced pattern, editing Ukraine–Russia–related content roughly half of the time. Only **Seryo93** and **UP9** focus predominantly on Ukraine–Russia topics, with nearly 60% of their edits targeting this domain.

IP users. In the case of IP users, the distribution differs noticeably from the pattern observed in Figure 3a. Five out of ten show a stronger focus on Ukraine–Russia–related articles, and three of these concentrate more than 80% of their edits on such topics. Overall, the values appear more extreme, indicating a more *all-or-nothing* pattern of behaviour compared with registered users. In other words, at least six IP addresses engage predominantly either in editing Ukraine- or Russia-related articles, or in editing non-related articles almost exclusively. In Figure 3a, the distribution is more balanced: only **Dbachmann**, **LSGH**, **Toddy1**, and possibly **Mazjac** exhibit a tendency to focus disproportionately on either Ukraine-/Russia-related content or on non-related content.

To complement this perspective, we also examine how weaponized and not weaponized edits are distributed within the same group of top contributors. The following figure displays the exact number of weaponized versus non-weaponized edits for each of the top ten users within the small database.

(a) Temporal distribution of *Weaponized* and *Not Weaponized* edits for the top ten registered users. [small db]

(b) Temporal distribution of *Weaponized* and *Not Weaponized* edits for the top ten IP users. [small db]

Figure 4: Comparison of the temporal distribution of *Weaponized* and *Not Weaponized* edits among the most active registered and IP users.

Interestingly—and consistent with the observations from Figure 3a—we see that **Dbachmann** made very few edits in 2006 and 2011 but contributed more than 100 edits in 2014. Approximately 33% of these were classified as weaponizing by the LLM. This is a non-negligible proportion for a user whose editing activity is otherwise almost entirely outside the Ukraine–Russia domain (about 99%). Moreover, **Dbachmann**, only contributed to the following articles : Crimea, History of Crimea, History of Ukraine and Russian annexation of Crimea, all of them classified as *Contentious Topics.* (see the Wikipedia Policy for more details 5.2.5) A user in a similar category was **LSGH**. However, the behavior observed in data differs: **LSGH** made around 150 edits in 2020 and more than 200 in 2021, with almost none of these being classified as weaponizing. A comparable contrast can be seen in another category discussed earlier. **Seryo93** and **UP9** were previously identified as users whose edits focus predominantly on Ukraine–Russia topics, with nearly 60% of all their edited articles falling into that category. Yet their weaponization profiles differ markedly. **UP9** has very few edits classified as *weaponizing*, whereas for **Seryo93**, approximately 38% of its edits in 2014 are labeled as weaponizing by the LLM.

Note. It can be interesting to correlate these patterns seen for the top ten registered users with their motivations. A former analysis for this purpose is done in Subsection 3.3.3

Regarding IP users, one can take a look to three interesting profiles here : user **77.252.47.155** and user **24.183.100.38** show no traces of weaponization while both are editing more than 70% Ukraine- or Russia-

related articles. User **217.118.81.18** did more *weaponized* edits than *Not weaponized*, yet he edited more than 70% non Ukraine or Russia-related articles. We can see an overall distinction between the two user types but there is also a large variability of profiles within each type.

Additional features related to user's contribution are also examined for deeper analysis, such as users' editing activity (total number of edits, edits per day or per month, whether their edits have been reverted, whether they revert others' edits, burstiness in editing patterns, etc.), namespace distribution and edit size. We conducted this analysis for all of the unique user's from the small database (6,922 users). In particular, burstiness may help determine whether weaponizing users engage in short, intense editing campaigns. We used Barabási burstiness index, defined as $B = \frac{(\sigma-\mu)}{(\sigma+\mu)}$, where μ = mean inter-event time, σ = standard deviation of inter-event time.

Using the `ucprop=editdiff` parameter, we found that the edits produced by the top ten registered users are, on average, approximately 1.5 times larger (in number of characters) than those made by the top ten IP-address users.

the ORES score[3] is the output of a machine learning model that assesses both the behavioural intent of an edit and its quality with respect to encyclopedic standards. In particular, it evaluates whether an edit was made to damage an article or was written in good faith. These models are therefore behavioural in nature. In addition, the `draftquality` score estimates whether a given revision appears acceptable as encyclopedic content.

All these features can then be compared with the number of weaponizing edits attributed to each user, allowing us to assess whether they are relevant variables to include when constructing the training set. Details are available in the book.

3.3.2 Users metadata

This section shifts the focus to user-level metadata. We retrieved metadata through Wikipedia's API using the `action=query` endpoint with `list=users` and the parameters `ucprop=blockinfo|groups|editcount|registration|emailable|gender`. This procedure was applied to both small and large databases (2,336 users and 19,796 users respectively).

Using the `ucprop=gender` parameter, we observe a strong gender imbalance among users in the small database, with 336 users identified as male, 28 as female, and 997 with no declared gender. This distribution is consistent with the well-documented gender bias in Wikipedia editing activity.[4] Interestingly, despite their small representation, female users exhibit a higher proportion of *weaponizing* edits than both male users and users with undeclared gender, a pattern that is observed consistently across both databases.

The parameters `ucprop=editcount` and `ucprop=registration` were used to examine whether weaponizing behaviour correlates with overall edit volume, account age, or a combination of both. These features provide insight into whether weaponization is associated with user experience or long-term engagement on the platform.

[3] `https://www.mediawiki.org/wiki/ORES/sv`

[4] `https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia`

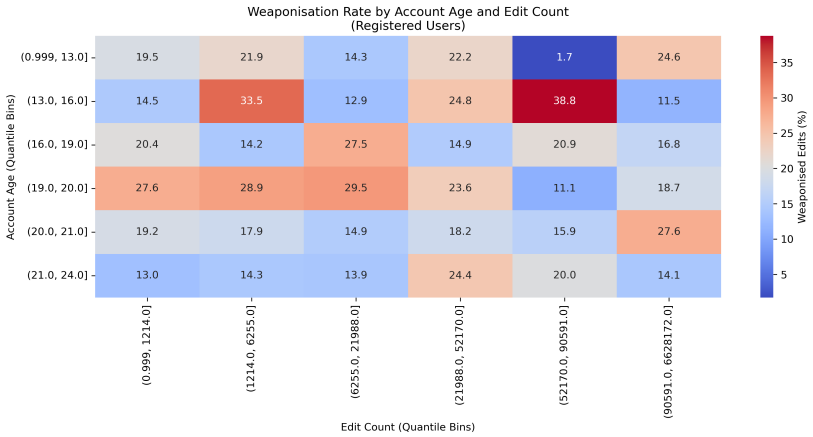


Figure 5: Heatmap of weaponization rates for registered users across quantile-based bins of account age and edit count. The color scale represents the mean proportion of edits classified as weaponizing (in percent) within each bin. [large db]

Finally, the `ucprop=blockinfo` parameter indicates that users who have previously been blocked are, in proportional terms, slightly more likely to produce weaponizing edits than unblocked users. More than 30 % of blocked users have less than 10 edits. 2% of them are female, 13% of male and the rest is unknown. 50% of them have been blocked within the two years after they registered and 25 % in less than a year.

It is important to note that the `ucprop=groups` parameter only reflects a user's current group memberships and does not capture historical group affiliations at the time an edit was made. Although logs tracking group changes do exist, exploiting this information at scale would require complex and computationally demanding processing, and was therefore not pursued in this study.

Note 1. Details for each `ucprop` are in the book.

Note 2. Metadata for IP users is not available, as these users do not possess registered accounts. Consequently, attributes such as the account registration date cannot be retrieved.

3.3.3 Users motivation

To gain preliminary insight into the *"why"* dimension of weaponizing behaviour, we conducted a qualitative, manual analysis as well as a computational, automatic analysis of the top ten subset active users. The objective was not to infer intent or assign definitive motivations, but to explore whether publicly available self-descriptions and interaction patterns might provide contextual signals relevant to understanding weaponisation dynamics.

The analysis relied primarily on information available on users' public *User pages* and *User talk pages*. Particular attention was paid to the presence of banners, templates, declared affiliations, comments, language usage, and recurring semantic patterns that may reflect ideological positioning, identity claims, or long-term engagement with specific narratives. Importantly, these elements can only offer indirect and partial insights. They do not allow, and should not be used for, determining an individual's cultural background, political beliefs, or intentions beyond what users explicitly choose to disclose. Insights are available in the book. Conclusions about this part is that we identify a large range of user's "personality", from the most patriotic and conservative ones to the most neutral one. Details are in the book.

3.3.4 Co-edited articles graph analysis

The construction of a co-edited articles graph is motivated by prior work on large-scale structural analyses of Wikipedia. In particular, previous visualisations[5] of the English Wikipedia have shown that representing articles and their relationships as a graph can reveal coherent thematic clusters through community-detection algorithms.

Building on this idea, we adapted a graph-based approach to identify structural patterns among *editors* that are co-editing the same article. Specifically, we constructed a graph in which nodes represent individual users and edges indicate that two users have edited at least one article in common. The edge weight corresponds to the number of articles they edited in common. This representation enables the detection of tightly connected groups of editors, which may reflect shared topical focus, coordinated editing behavior, or the formation of echo chambers.

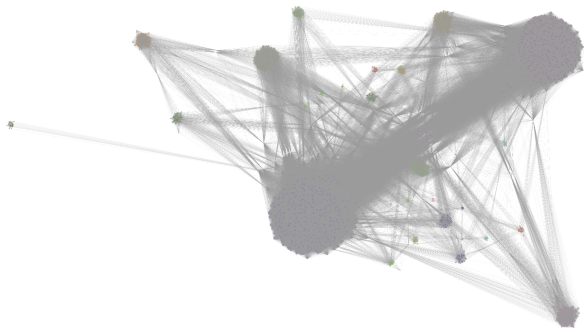


Figure 6: Leiden graph with edge weights filtered to 1 and above, visualized with Gephi software. It contains **2,243 nodes**, **287,323 edges** and **16 clusters**.

The graphs were constructed using the `python-igraph` library, and the resulting network was visualized and explored using the Gephi software. The community-detection methods applied is the Leiden algorithm to identify clusters of editors whose editing activity overlaps significantly.

We decided to show two very different plots. The first one is without any edge weights filtering. Thus, we will have every links between two users displayed, even if they only edited a single article in common. The graph displayed by the software showed **2,243 nodes**, **287,323 edges** and **16 clusters**. This graph, even if the weights were not filtered, showed a really interesting visualization, were some echo chambers are visible.

By looking to Figure 6, we can clearly identify the first two main clusters (cluster 0 and cluster 1). Things get interesting when we look to the outliers. A single, very lonely cluster (cluster 11), on the left side of the image, is barely identifiable. This cluster contains 23.5% of weaponizing edits, which make it one of the most weaponizing clusters among the 16 clusters. Interestingly, it contains 32 nodes, where 20 of them are IP users. They only focused on editing a single articles, *Crimean Tatars*, and all edits (46 in total) were done within a year, from December 2004 to December 2025.

The same analysis can also be done for cluster 6 and 8. Details are available in the book.

3.3.5 IP address geolocalisation

Because unregistered users can only be identified through their IP addresses, an IP geolocation service can be used to estimate their approximate locations. By querying the API provided by such a service, it is possible

[5] `https://www.youtube.com/watch?v=JheGL6uSF-4`

to visualise these locations on a world map. The resulting distribution reveals a strong concentration of unregistered editors in English-speaking countries, particularly the United States and the United Kingdom.

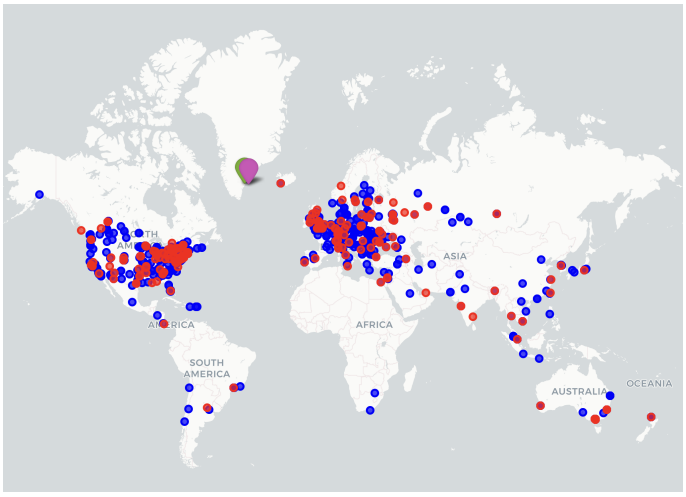


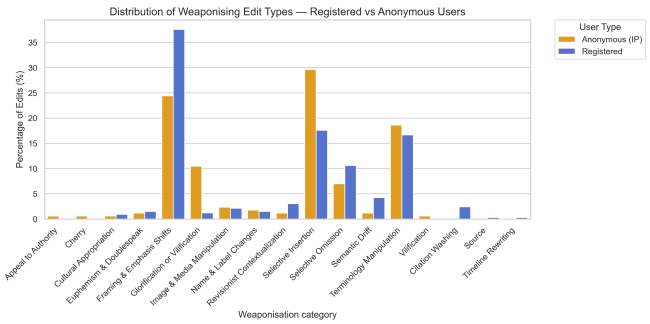
Figure 7: IP Localization Map. red dots are *weaponizing* IP user and blue dots are *Not weaponizing* IP user. Green and pink pins are the average localisation. (green for *Not weaponizing* and pink for *weaponizing*

4 Weaponizing Users — Fine-Grained Analysis

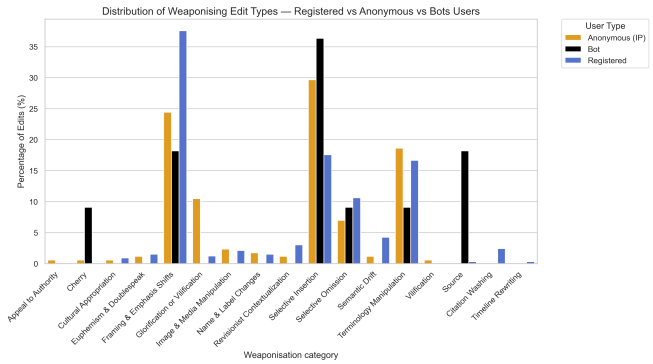
4.1 Context

This section focuses exclusively on edits classified as *Weaponized* by the LLM. Restricting the analysis to this subset allows for a more detailed examination of manipulation strategies. In addition, we leverage the fine-grained taxonomy developed in Hidi's project [6], which categorizes weaponizing edits according to the semantic and structural techniques employed.

4.2 Taxonomy Distribution by User Type



(a) Distribution of weaponizing taxonomy categories by user type.



(b) Distribution of weaponizing taxonomy categories across user types, including bots.

Figure 8: Comparison of the proportion of Ukraine–Russia–related edits among the top ten registered users and the top ten IP users for the fine grained database.

Figure 8a shows the distribution of weaponizing taxonomy categories across registered and anonymous (IP) users. Clear differences emerge between these two groups. Registered users predominantly engage in cat-

egories associated with **Framing and Emphasis Shifts**, which typically involve longer edits that subtly recontextualise or rephrase existing content. In contrast, anonymous users are more frequently associated with **Selective Insertion**, characterized by shorter and more targeted modifications.

Extending this comparison to include bot accounts (Figure 8b) reveals a distinct behavioral profile. Bot-generated edits are concentrated in a limited number of taxonomy categories, most notably **Selective Insertion**, as well as **Cherry-Picking** and **Source-Biasing**. These patterns are consistent with automated editing processes, which often involve systematic content insertion or source replacement rather than extensive textual rewriting.

We next investigated differences in edit magnitude, measured as the absolute length of edits, across the various user types. The resulting visualisation highlights distinct behavioral patterns. We notice that the size of the edits strongly differs across different categories regarding the two user types.

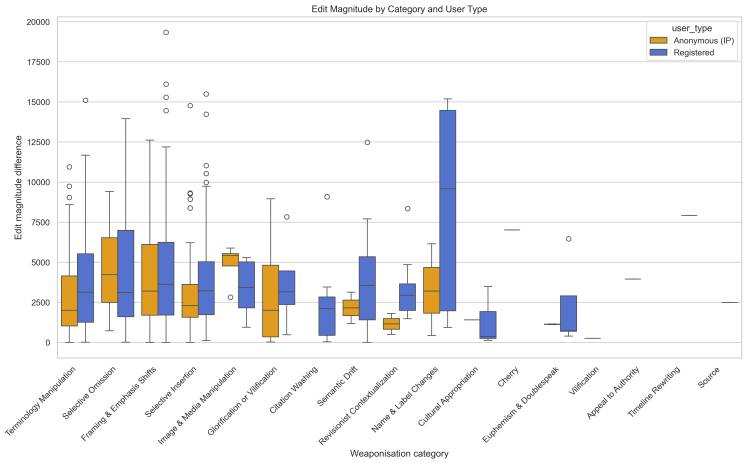


Figure 9: Differences of edit magnitudes for each user type with respect to the taxonomy

5 Article Policy Analysis

5.1 Context

Carefully studying Wikipedia's policies was essential to this project for several reasons. First, it provides a clearer understanding of how Wikipedia functions as a socio-technical system, including its editorial structure, governance mechanisms, and policy development processes. Second, it offers insight into user roles, group memberships, and permission hierarchies, helping to contextualise editing behaviour over time. Finally, this analysis reveals how both article content and article importance are formally assessed within Wikipedia, and how these classification mechanisms can themselves become targets of manipulation. In particular, the article assessment system gives rise to a distinct form of *structural weaponization*, in which the textual content of the article is not the primary target, but rather its quality assessment and importance status as perceived by editors and readers.

5.2 Wikipedia's policy in a nutshell

We will introduce here the different Wikipedia's policies concept that are next used for the articles' database analysis.

5.2.1 User Hierarchy

Types	How They Get Status	Main Rights & Abilities	Position in Hierarchy
Unregistered Users (IP editors)	No account	Edit most pages (except protected pages); create talk pages; cannot upload media	Lowest rights but considered equal contributors
Registered Users	Create an account	Edit pages; create new pages; upload media; hide their IP address	Baseline registered users
Autoconfirmed Users	Automatic after 4 days	Additional privileges (move pages; edit semi-protected pages)	Slightly above registered users
Extended Confirmed Users	30+ days and 500+ edits	Revert changes; delete pages; edit templates	Mid-level tier
Administrators	Community approval + bureaucrat action	Block editors; protect/unprotect pages; delete/restore pages	High technical authority
Bureaucrats	Elected volunteers	Promote admins; assign bot flags; remove admin rights	Above administrators
ArbCom	Elected annually	Resolve conduct disputes; impose sanctions	Supreme dispute-resolution body
Stewards	Globally elected	Assign/remove permissions globally; check-user + oversight rights	Global technical authority

Table 2: Overview of Wikipedia User Groups and Authorities

5.2.2 Vital Articles

Within the English Wikipedia, a subset of articles is designated as Vital Articles[6]. This initiative identifies the most important topics for the encyclopedia and serves both as a prioritisation framework for editorial improvement and as a centralised watchlist for highly visible content. Vital Articles are organised into five hierarchical levels: Level 1 includes the ten most essential articles, Level 2 expands this set to one hundred articles, and each subsequent level progressively broadens the scope by building on the previous one.

Vital article level	Target number of vital articles	Current situation (September 8, 2025)		
		Current number	At or above -class (%)	At or above class (%)
1	10	10	10 (100%)	6 (60%)
2	100	100	88 (88%)	30 (30%)
3	1,000	1,000	699 (70%)	222 (22%)
4	10,000	10,002	*	*
5	50,000	50,014	*	*

Figure 10: Distribution of articles across Vital Article levels.

[6] https://en.wikipedia.org/wiki/Wikipedia:Vital_articles

For example, Level 1 comprises ten foundational articles, including *The Arts*, *Earth*, *Human*, *Human history*, *Life*, *Mathematics*, *Philosophy*, *Science*, *Society*, and *History*.

5.2.3 WikiProject

WikiProjects[7] are collaborative groups of contributors organised around specific topic areas (e.g., mathematics, Kenya, greek mythology, etc...) with the goal of improving and maintaining related articles. Articles may belong to one or multiple WikiProjects, which enables topic-specific monitoring and coordinated editorial activity.

Within each WikiProject, articles are assigned an *importance* rating. This rating tries to capture their relative significance to readers. This grading system addresses the question: *"How likely is it that a typical Wikipedia reader will need this article?"* The importance scale provides a structured mechanism for prioritising editorial attention and resource allocation within a project.

5.2.4 Article Quality and Importance Assessment

Wikipedia evaluates articles along two complementary dimensions: *quality* and *importance*. These assessments are assigned within the context of individual WikiProjects and reflect both the intrinsic completeness of an article and its perceived relevance within a given topic area.

Article quality is assessed using a letter-based scale that primarily measures factual completeness, with language quality and structural organisation also taken into account. The main quality grades are summarised below (see Wikipedia:Content assessment[8] for full documentation):

- **FA (Featured Article):** Professional, comprehensive, and authoritative.
- **FL (Featured List):** Professional standard, covering a well-defined scope.
- **A-Class:** Highly complete and useful; near-professional quality.
- **GA (Good Article):** Useful to most readers, with no major issues.
- **B-Class:** Substantial content, though not fully comprehensive.
- **C-Class:** Basic coverage, suitable for casual readers.
- **Start-Class:** Limited content requiring significant expansion.
- **Stub-Class:** Minimal content, often only a short introduction.

In parallel, article importance is assessed within individual WikiProjects according to how essential an article is to a given topic area. These two dimensions are jointly tracked by the Wikipedia Version 1.0 Editorial Team[9], which maintains a structured overview of article assessments across projects.

[7] `https://en.wikipedia.org/wiki/Wikipedia:WikiProject`
[8] `https://en.wikipedia.org/wiki/Wikipedia:Content_assessment`
[9] `https://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team`

Status	Template	Meaning of Status
Top	{{Top-Class}}	This article is of the utmost importance to this project, as it forms the basis of all information.
High	{{High-Class}}	This article is fairly important to this project, as it covers a general area of knowledge.
Mid	{{Mid-Class}}	This article is relatively important to this project, as it fills in some more specific knowledge of certain areas.
Low	{{Low-Class}}	This article is of little importance to this project, but it covers a highly specific area of knowledge or an obscure piece of trivia.
None	None	This article is of unknown importance to this project. It remains to be analyzed.

Figure 11: WikiProject importance scale and corresponding meaning of each level.

The Version 1.0 Bot stores quality and importance ratings for each article and WikiProject, enabling large-scale comparisons and longitudinal analysis. Figure 12 illustrates the distribution of articles by quality and importance for *WikiProject Ukraine*.

Ukraine articles by quality and importance

Quality	Importance						
	Top	High	Mid	Low	NA	???	Total
FA		4	3				7
FL		1	3	7			11
A		2	2	1			5
GA		11	24	30		1	66
B	35	138	136	251		160	720
C	49	254	382	1286		745	2716
Start	19	261	826	5056		2702	8864
Stub		24	399	6967		5714	13104
List	6	30	72	323	448	135	1014
Category					12262		12262
Disambig					88		88
File					357		357
Portal					1		1
Project					23		23
Redirect			1	140	751		892
Template					979		979
NA					3		3
Other				2	101		103
Assessed	109	725	1848	14063	15013	9457	41215
Unassessed						4	4
Total	109	725	1848	14063	15013	9461	41219

Last updated: 11/27/25, 2:27 AM

Figure 12: Distribution of articles by quality and importance within WikiProject Ukraine.

Importantly, the Version 1.0 infrastructure also allows comparisons across WikiProjects. By contrasting quality and importance ratings between projects (e.g., Ukraine- and Russia-related WikiProjects), it becomes possible to identify articles whose perceived relevance diverges across communities, potentially revealing patterns of *structural weaponization* that do not involve direct textual modification.

5.2.5 Contentious Topics

Some Wikipedia pages are assigned CT by the Arbitration Committee after users submit a CT request. Only extended-confirmed editors may make edits related to the topic area, though editors who are not extended-confirmed may post constructive comments and make edit requests related to articles within the topic area on article talk pages. One must highlight two things. First, by *Topic*, the policy means that any Wikipedia article related to that topic may be tagged by the *CT* keyword. Second, there is a key difference between *Controversial Topic* and *Contentious Topic*. The *controversial* keyword has no significant meaning in the Wikipedia Policy. Therefore, an article can be considered as *controversial* by a community

of users but may not be declared as *contentious*. Only *contentious* keyword is the one recognized by the Wikipedia Policy. Some non exhaustive *Contentious Topic* are Armenia, Azerbaijan, or related conflicts, the Arab–Israeli conflict, Abortion, Climate change, the Balkans or Eastern Europe, etc...

5.2.6 Protection Status

In some circumstances, pages may need to be protected from modification by certain groups of editors. Pages are protected when there is disruption that cannot be prevented through other means, such as blocks. Protection is a technical restriction applied only by administrators, although any user may request protection. Protection can be indefinite or expire after a specified time.

By looking to Wikipedia's Protection Policy[10], we can find a lot of different protection status that has been created for specific task. First, we have to understand that there are 4 different types of protection :

- **edit protection :** protects the page from being edited by certain user type
- **move protection :** protects the page from being moved or renamed
- **creation protection :** prevents a page from being created
- **upload protection :** prevents new versions of a file from being uploaded

The last two types of protection are less important, so we can focus primarily on the first two.

5.3 Blend everything together

Now that the foundational aspects of Wikipedia's organisational structure and article system have been established, the next step is to collect the variables listed above for each user or article. Once compiled, these metadata elements provide valuable insights into user behavior and broader patterns relevant to weaponizing culture heritage in Wikipedia. Article-level attributes—including *quality assessment*, *vital article* status, *contentious topic* designation, *WikiProject importance*, and *protection status*—were collected around 11 November. With the exception of protection status, these attributes are not visible in the article's Main namespace but are instead embedded as raw text within the Talk page namespace, where they are manually encoded by editors. Each feature within the wikitext is encoded as a template (e.g., `WikiProject banner shell|class=C|vital=yes|1=WikiProject Ukraine|importance=Top|crimea=yes`). No specific Wikipedia API endpoints were used to retrieve this information. Instead, the data were extracted directly from the Talk page content following the procedure described below:

1. For *Quality Assessment (QA)*, *WikiProject importance*, and whether an article is classified as *contentious*, we first retrieved the wikitext (the underlying markup of a page) from each article's Talk page using the API with `prop=revisions` and `rvprop=content`. We then applied a `parse_assessment` function to process the wikitext and extract the relevant attributes. **Note:** As described in subsection 5.2.3, an article may belong to multiple WikiProjects, each assigning its own importance rating (Top, High, Mid, Low, or None). In this study, the stored importance value corresponds to *WikiProject Ukraine*. In the rare cases where an article is not associated with WikiProject Ukraine, the recorded importance rating corresponds to the first importance level appearing in the first WikiProject listed in the wikitext.
2. The *Vital Article level* (from 1 to 5, or None; see subsection 5.2.2) was more difficult to extract. This information required direct HTML scraping of the article's webpage using *BeautifulSoup*, a Python library specialised in web scraping.

[10] `https://en.wikipedia.org/wiki/Wikipedia:Protection_policy`

3. Because the *protection status* of a page is directly available in the Main namespace, it was retrieved via the API using `action=query`, with the parameters `prop=info` and `inprop=protection`. This allowed us to obtain the current protection level for each article.

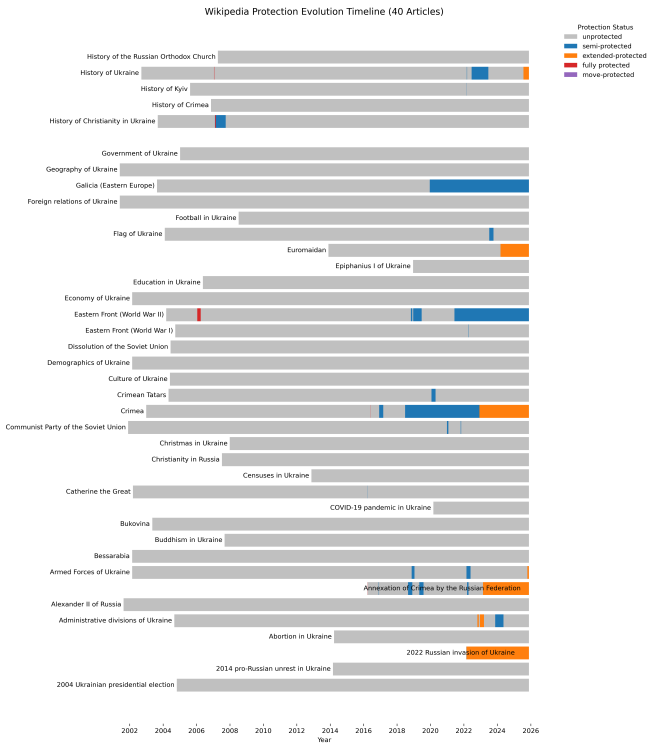
In addition to the static analysis presented above—where each feature reflects the policy state of an article as of 11 November 2025—we also conducted a longitudinal analysis of the evolution of each feature. Unlike other policy attributes, protection-level changes are systematically logged by Wikipedia, making it possible to reconstruct how an article's protection status has evolved over time using the Wikimedia API's `action=query` endpoint with `list=logevents` and `letype=protect`. Below, we outline the full pipeline used to retrieve and analyse all other features, including vital level, quality rating, and importance status:

1. **Retrieval of templates from Talk pages.** For each article, we retrieved the full revision history of its Talk page using the Wikipedia API. Each revision was then parsed to identify templates encoding changes in article-level attributes. All detected changes for a given article were stored in a chronological timeline.

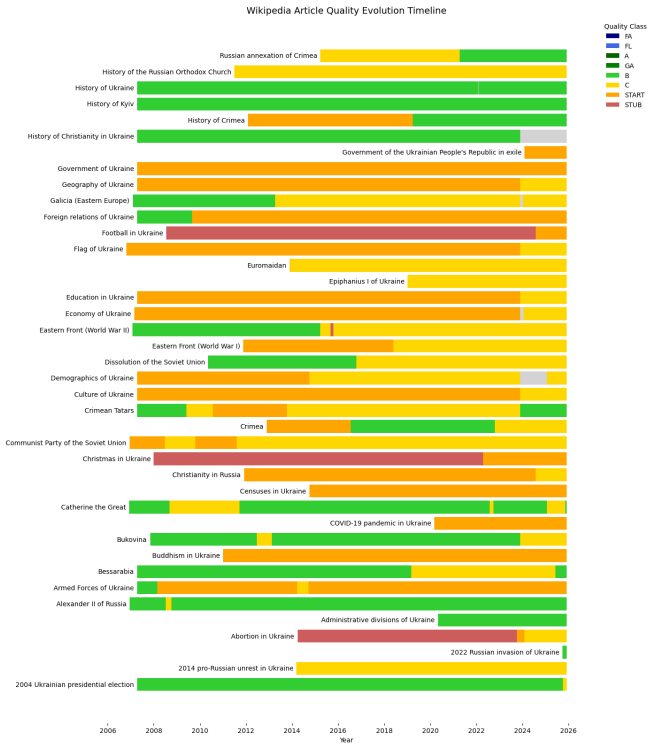
2. **Construction of attribute-specific dataframes.** After cleaning the timelines, we first visualised the temporal evolution of each attribute using Gantt-style plots (see Figure 13). Each timeline was then converted into a dataframe storing, for each article, the time at which an attribute was logged and the value to which it was changed.

3. **Conversion into a unified dataset.** Finally, the reconstructed dataframes were merged into the preprocessed large database, allowing each edit to be associated with the protection level, importance status, quality grade, and vital level in effect at the time it was made. This step enables temporal analyses linking user behaviour and weaponising edits.

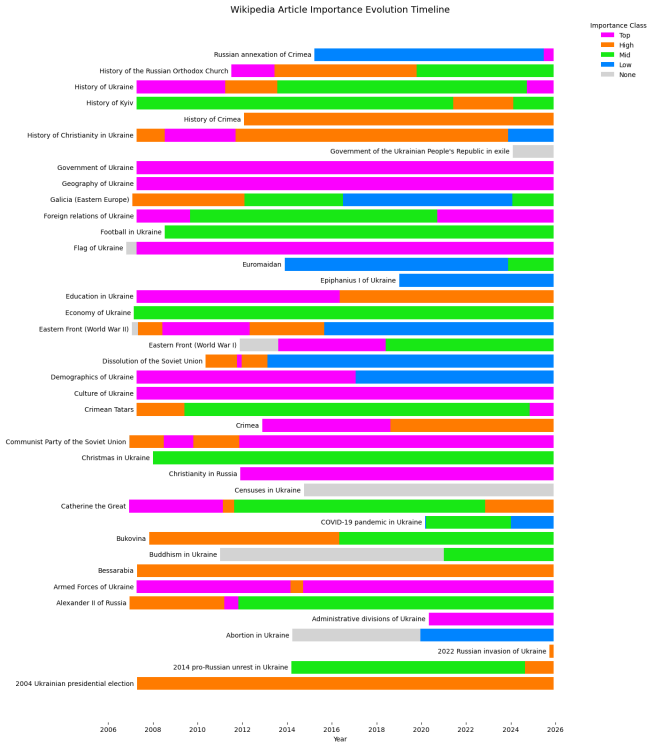
The final four plots show the distribution of *weaponising* versus *non-weaponising* edits with respect to the different features at the time each edit was made.



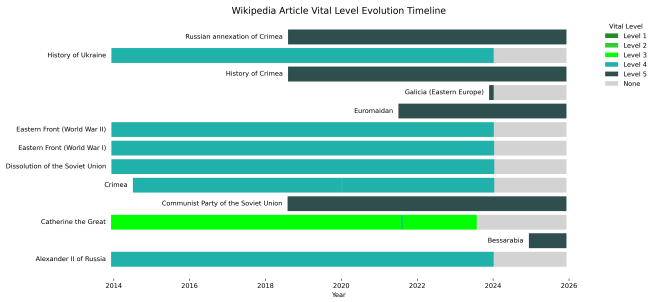
(a) Evolution of protection-level changes



(b) Evolution of quality grade changes



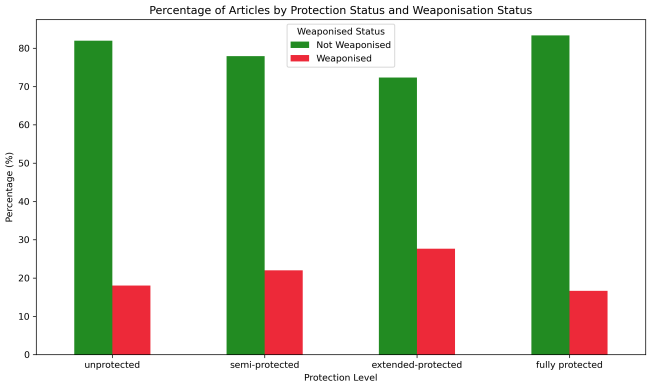
(c) Evolution of importance status changes



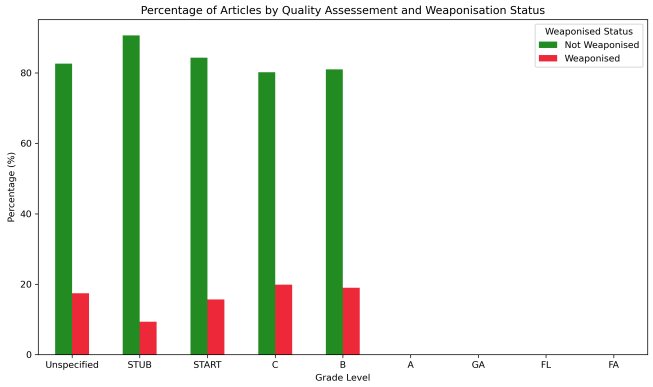
(d) Evolution of vital level changes

Figure 13: Gantt-style timelines showing the evolution of policy-related attributes across 40 Wikipedia articles

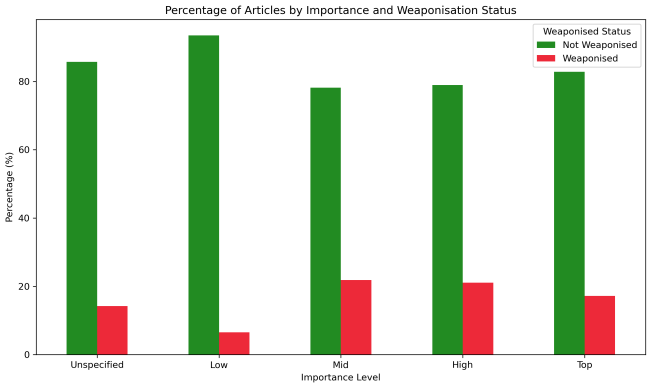
The final four graphs visually show the evolution of the different features over time.



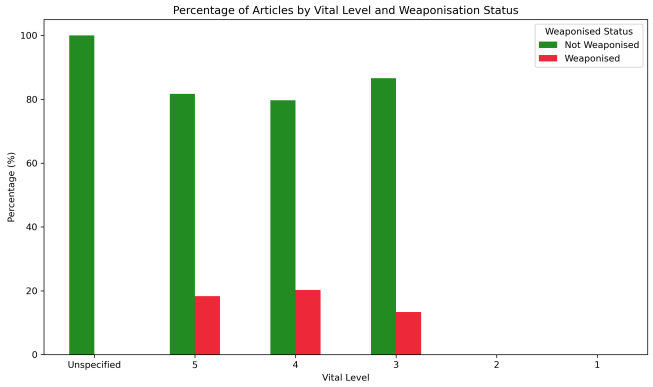
(a) Evolution of protection-level changes



(b) Evolution of quality grade changes



(c) Evolution of importance status changes



(d) Evolution of vital level changes

Figure 14: Panel that shows how weaponized edits are distributed regarding the 4 features identified through Policy Analysis for the 40 articles.

The absence of a clear correlation between weaponization and some specific value is frustrating. One explanation that can be raised it that the article policy is really niche and unknown for many users.

6 Conclusions and Future Directions

After exploring the data, correlating user types with multiple features, analysing the visualisations, and conducting an in-depth examination of user metadata, contributions, and behavioural patterns, one central observation clearly emerges: weaponization is far more complex and subtle than it may initially appear. It cannot be reduced solely to vandalism, propaganda, or straightforward disinformation. Instead, the analyses conducted throughout this work reveal a rich and complex phenomenon, in which manipulative practices are often embedded in legitimate-looking editorial actions.

We saw that this complexity is reflected in the empirical results. Plots distributions are frequently well balanced and graphical representations rarely point unambiguously toward a single, clear-cut interpretation. This pattern makes the interpretations and the analysis more difficult. Interestingly, we have seen also some clear differences between the registered and the anonymous user types, especially in the users contribution and weaponized taxonomy.

A major contribution of this project lies in the extensive analysis of Wikipedia's governance system and editorial policies, that took more than 8 weeks. This investigation made it possible to identify and confirm a structural form of weaponization that does not primarily operate through semantic manipulation. Instead, it manifests through the attribution (and subsequent modification) of article-level assessments, such as

quality grades, importance ratings, vital level, and protection status. These mechanisms shape the visibility, perceived legitimacy, and editorial prioritisation of articles, and thus represent a powerful yet much less visible and well known lever for influencing collective knowledge production. The absence of a clear correlation between semantic and structural weaponization can largely be explained by the niche nature of article-level policy features, which may not necessarily be a primary targets of semantic weaponization efforts.

6.1 Control Analysis

t would be both necessary and informative to compare the results of this analysis—particularly the contrast between registered and unregistered user types—against a control group. To this end, a complete and ready-to-run pipeline was developed in parallel using a set of 40 articles related to the topic of music. By strictly replicating the pipeline employed in Hidi's work, we collected 40 articles and more than 200,000 combined edit diffs, which are ready to be classified via the ChatGPT API as either Weaponized or Not Weaponized.

The topic *Music* was selected somewhat arbitrarily as the articles inside of it.

Page	**Edits**	**Page**	**Edits**
Alan Clark	894	Pink Floyd	18111
James Brown	8480	Daft Punk	8954
Music theory	3383	Jazz	11820
Guitar	6906	Dire Straits	4131
Music of Africa	2733	Rock and roll	8832
Red Hot Chili Peppers	21645	Piano	7927
Eric Clapton	10102	Country music	8486
Major scale	918	Stevie Wonder	5822
Royal Albert Hall	2112	Swing	406
Major chord	199	John Frusciante	7306
Pop music	9681	Bob Marley	7142
Minor chord	199	Marshall Amplification	1483
Wolfgang Amadeus Mozart	8547	Ludwig van Beethoven	9109
Nirvana (band)	13978	Jimi Hendrix	14167
Joseph Haydn	5337	Saxophone	6499
Paul Kalkbrenner	367	Electronic music	4393
Classical music	5607	Nina Simone	3461
Rolling Stone	4161	Fender	84
Trumpet	7430	BTS	8480
Jean-Jacques Goldman	530	Funk rock	911

We count a total of 250,733 edits in 40 articles.

6.1.1 Data Collection

Because it's a control group, I carefully reproduce Hidi's method. For every selected article title, I leveraged the Wikimedia API's `action=query` endpoint with prop=revisions and the parameters `rvprop=timestamp|user|comment|content` to stream the entire revision history. Continuation tokens were handled automatically to ensure no edits were missed. I reconstructed each article's state in chronological order and then computed unified diffs between successive versions via Python's difflib.unified diff, capturing additions, deletions, and context lines. Each edit was stored as a JSONL record containing the pre-edit text, the post-edit text, the diff hunks, and associated metadata.

Here is an example of output for the article "Bob Marley" :

```
root  7129 items
   0
   version      "first_version"
   Content      "Robert Nesta '''Marley''' a.k.a Berhan '''Selassie''' was a
                 "King of Reggae music" came from the ghettos of [[Jamaica]],
```

playing, teaching and singing for a long period in 70’s and 80’s.
His best album made with reggae group [[The Wailers]] in [[London]]
was perhaps [[Exodus]]. His work stayed unstained all the way to
his last original album [[Upriseing]] because all he ever had was
songs of freedom and it will live forever. His life is a synonyme
for a struggle of any kind (especially in his native island and in
African continent) and for the opressed even today. Some argue he
was trully "boboshanti solder" in this "Babylon system" but perhaps
they are all from the group of "Babylonians".
=== General Marley’s quotations and lyrics ===
"I will talk no more, cause this is not a joke" [[Bob Marley|Robert
Nesta Marley]]
"Dema tired see me face, Dema cant get mi outa di race"
[[Bob Marley|Robert Nesta Marley]]
"Jamaica is small, Ethiopia is an adventure..."
=== Quotation of his life and work from others ===
’’’See also:’’’ [[Reggae]], [[Roots Rock Reggae]], [[Rastafarianism]],
[[Ras Tafari]], [[Jamaican English]], [[Amharic]], [[Ethiopia]].

1
version "diff"
Timestamp "2002-02-27T07:54:04Z"
User "XJaM"
Comment "Just to start from scrach to memorate this great mon."
Diff "Initial revision"

2
version "diff"
Timestamp "2002-03-12T15:27:32Z"
User "Koyaanis Qatsi"
Comment "*"
Diff "--= +++ @@ -1,17 +1,3 @@
-Robert Nesta ’’’Marley’’’ - -a.k.a Berhan ’’’Selassie’’’ was a
"King of Reggae music" came from the ghettos of [[Jamaica]], playin
g, teaching and singing for a long period in 70’s and 80’s; Marley
stayed unstained all the way to his last original album [[Upriseing]]
because all he ever had was songs of freedom and it will live forever.
His life is a synonyme for a struggle of any kind (especially in his
native island and in African continent) and for the opressed even toda
y this "Babylon system" but perhaps they are all from the group of
"Babylonians’.
=== General Marley’s quotations and lyrics ===
"I will talk no more, cause t

6.1.2 Weaponization Detection : Prompt Design and LLM Classification

Citing Hidi’s report :

> *“Our classification pipeline begins by framing each Wikipedia edit as a structured JSON record and then asking a Large Language Model to reason over the exact changes. We serialize either the full ‘first version’ text or a ‘diff’ record (which includes the ISO timestamp, user identifier,*

> *revision comment, and unified diff hunk) into a JSON snippet. The model is instructed to restate the change in natural language, and then issue exactly one of the labels* `Weaponized` *or* `Not Weaponized`*, followed by a brief justification. This two-step instruction mimics a persona-based approach by eliciting an explicit summary of the edit before the binary judgment, yet requires no in-prompt exemplars."*

(Hidi, 2025)

6.2 Limitations.

6.3 Limitations

This study is subject to several limitations. First, the classification of edits as weaponized or not weaponized relies on labels produced by a large language model and therefore does not constitute ground truth; only expert human annotation could provide definitive validation. Second, many user- and article-level features are available only in their current state, making it difficult to accurately reconstruct the historical context in which past edits were made. In addition, Wikipedia's policies, quality assessments, and importance ratings evolve over time, which further complicates longitudinal analysis. Finally, the scope of the study is limited to a predefined set of articles and to the English Wikipedia, which may restrict the generalisability of the findings to other languages, topics, or cultural contexts.

A Extended Keyword Lists and Plots

The following ChatGPT (GPT 5.0) prompt was used to generate the keywords lists.

Note. The prompt wasn't asked in a fresh new chat but was generated in a ChatGPT project, already filled with contexts and other related conversations :

```
Using the Wikipedia API allcontribs parameter, I retrieved for a list of 10 users all of their e
```

This annex contains the small and extended list of keywords used to classify Ukraine–Russia–related articles.

```
ukraine_keywords_small = [
    "ukraine", "ukrainian", "kyiv", "kiev", "crimea", "crimean", "kuban",
    "donbas", "donetsk", "luhansk",
    "maidan", "yanukovych", "yushchenko", "zelenskyy", "poroshenko",
    "catherine", "bukovina", "bessarabia", "eastern",
    "euromaidan", "dnipro", "odessa", "sevastopol", "putin", "rus'",
    "russia", "russian", "moscow", "kremlin", "soviet"
]

ukraine_keywords_large = [
# core country and people
"ukraine", "ukrainian", "kyiv", "kiev", "crimea", "crimean", "kuban", "donbas", "donetsk", "luhansk"
"maidan", "yanukovych", "yushchenko", "zelenskyy", "poroshenko",
"catherine", "bukovina", "bessarabia",
"eastern", "euromaidan", "dnipro", "odessa", "sevastopol", "putin",
"rus'", "russia", "russian", "moscow",
"kremlin", "soviet",
# politics & government
"verkhovna rada", "president", "prime minister", "parliament",
"government", "cabinet", "federation",
"referendum", "annexation", "independence", "revolution", "reforms",
"corruption", "sanctions",
"occupation", "treaty", "agreement", "ceasefire", "negotiations",
"elections", "coup", "unification",
# geography & regions
"zaporizhzhia", "mariupol", "kharkiv", "kherson", "mykolaiv",
"chernihiv", "sumy", "poltava", "vinnytsia",
"lviv", "ivano-frankivsk", "ternopil", "lutsk", "uzhhorod",
"dnipropetrovsk", "donetsk oblast",
"luhansk oblast", "transcarpathia", "prykarpattia", "galicia",
"novorossiya", "black sea", "azov sea",
# historical references
"kyivan rus", "tsar", "imperial", "empire", "ussr", "communist",
"lenin", "stalin", "bolshevik",
"cold war", "perestroika", "glasnost", "collapse", "partition",
"catherine the great", "brezhnev",
"chernobyl", "orange revolution", "revolution of dignity",
"holodomor", "soviet union", # war and military
"invasion", "occupation", "annexed", "frontline", "offensive",
"defense", "army", "forces", "military",
```

```
"russian troops", "ukrainian forces", "separatist", "rebels",
"paramilitary", "nato", "eu", "un", "war",
"conflict", "shelling", "bombing", "airstrike", "occupation forces",
"mobilization", "martial law",
# culture, identity & language
"language", "identity", "heritage", "culture", "orthodox", "church",
"patriarch", "ukrainian language",
"russian language", "minority", "bilingual", "autonomy",
"nationalism", "independence day", "flag",
"anthem", "symbol", "national identity", "sovereignty",
# current / modern references
"donbas war", "russian invasion", "ukrainian front", "crimea bridge",
"moskva cruiser", "ukrainian army",
"russian army", "zelensky", "kremlin propaganda", "occupation
administration", "territorial defense",
"european union", "eu membership", "nato membership", "nato
expansion", "eu sanctions", "ukraine war",
"full-scale invasion", "special military operation", "mobilisation", "referendum in crimea",
# other
"gas pipeline", "north stream", "energy crisis", "grain corridor",
"black sea fleet", "peace talks",
"donetsk people's republic", "luhansk people's republic", "kyiv
oblast", "liberation", "resistance",
"occupation zone", "ukrainian refugees", "mariupol steel plant",
"azovstal", "bucha", "irpin", "kharkiv offensive"
]
```

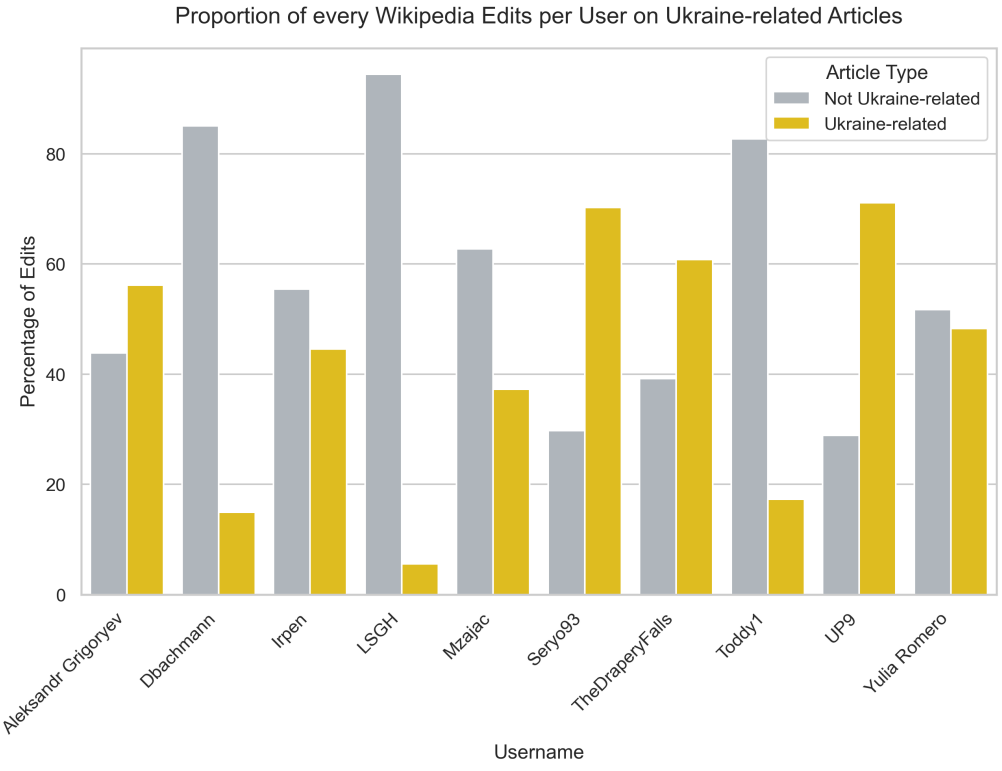


Figure 15

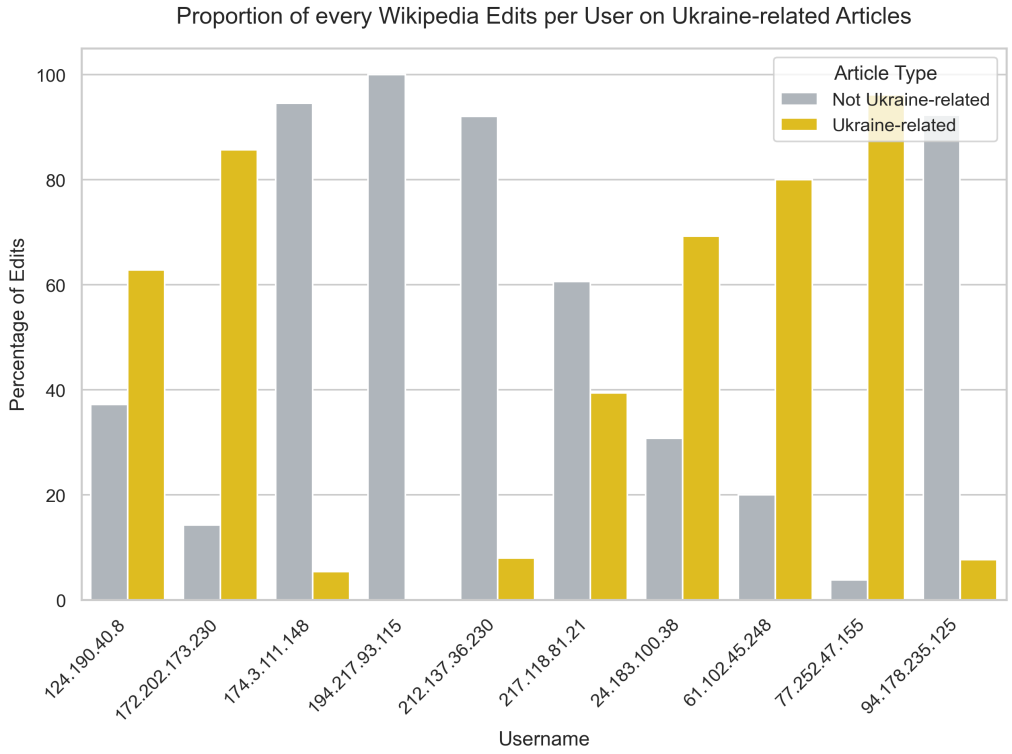


Figure 16

B Additional Tables

Metric	Value
Total edits	69908
Total unique users	19769
Number of unique articles	40
Bot users	236
Anonymous users	10609
Registered users	8924
Bot edits	3910
Anonymous edits	17820
Registered edits	48178
weaponized edits	13445
Non-weaponized edits	56463

Table 3: Summary statistics extracted from the dataset.

C Weaponization Taxonomy Definitions

This subsection provides concise definitions of the weaponization categories used throughout the fine-grained analysis. The taxonomy was originally developed in the semester project by Hidi et al. [6] and is reproduced here to ensure clarity, reproducibility, and consistency for future analyses and model development.

article	Not weaponized Edits	weaponized Edits	Total Edits
Eastern Front (World War II)	5651	1398	7049
Catherine the Great	5871	1035	6906
2014 pro-Russian unrest in Ukraine	3898	1275	5173
Annexation of Crimea by the Russian Federation	2763	1468	4231
Dissolution of the Soviet Union	3167	595	3762
Crimea	2508	1132	3640
Alexander II of Russia	2634	415	3049
Eastern Front (World War I)	2211	403	2614
Crimean Tatars	1937	533	2470
Economy of Ukraine	1951	338	2289
Armed Forces of Ukraine	1919	364	2283
Galicia (Eastern Europe)	1943	311	2254
History of Ukraine	1724	520	2244
Demographics of Ukraine	1771	230	2001
Euromaidan	1463	478	1941
Bukovina	1468	395	1863
COVID-19 pandemic in Ukraine	1588	110	1698
Bessarabia	1166	300	1466
Communist Party of the Soviet Union	1199	123	1322
Flag of Ukraine	1130	139	1269
History of Christianity in Ukraine	834	317	1151
Administrative divisions of Ukraine	833	190	1023
History of Crimea	726	254	980
History of Kyiv	696	233	929
Culture of Ukraine	786	116	902
2004 Ukrainian presidential election	660	89	749
Foreign relations of Ukraine	606	133	739
Government of Ukraine	624	64	688
Education in Ukraine	499	62	561
Christmas in Ukraine	433	90	523
Geography of Ukraine	455	49	504
History of the Russian Orthodox Church	363	125	488
Epiphanius I of Ukraine	307	45	352
Christianity in Russia	236	61	297
Football in Ukraine	242	27	269
Censuses in Ukraine	58	7	65
Buddhism in Ukraine	55	2	57
Abortion in Ukraine	44	3	47
Government of the Ukrainian People s Republic in exile	37	9	46
2022 Russian invasion of Ukraine	7	7	14
Total	56463	13445	69908

Table 4: Summary statistics extracted from extended the dataset.

Each category corresponds to a distinct manipulation strategy observed in Wikipedia edits related to cultural heritage and geopolitical narratives.

- **Framing and Emphasis Shifts** Edits that subtly reframe existing content by modifying wording, emphasis, or contextual placement without necessarily introducing new factual claims. This includes changes in tone, ordering of information, or selective contextualisation that promotes a particular interpretation.
- **Selective Insertion** Short and targeted edits that insert specific facts, phrases, or statements while leaving the surrounding content largely unchanged. These interventions often aim to introduce a particular narrative element with minimal textual modification.
- **Cherry-Picking** The selective inclusion of data points, statistics, or historical facts that support a specific viewpoint, while omitting relevant counter-evidence or alternative perspectives. This strategy can distort interpretation without altering factual accuracy at the sentence level.
- **Source-Biasing** Edits that manipulate the perceived credibility of content by adding, removing, or replacing citations. This includes substituting reliable sources with partisan or low-quality references, or selectively citing sources that favour a specific narrative.
- **Structural Manipulation** Changes that alter the structural organisation of an article—such as section headers, ordering, or segmentation—in ways that affect how information is accessed or prioritised by readers.
- **Content Removal or Erasure** The deletion of text, references, or entire sections that previously conveyed relevant cultural, historical, or political information, potentially leading to the suppression of specific narratives or perspectives.

These categories are not mutually exclusive: a single edit may exhibit characteristics of multiple weaponization types. In the present dataset, each edit is assigned a primary category based on its dominant manipulation strategy. The taxonomy can therefore be used both as a multi-class label space and as a source of higher-level semantic features for downstream modeling tasks.

D Edits per article distribution per dataset

For the **small dataset :**

Articles	Not Weaponised Edits	Weaponised Edits	Total Edits
COVID-19 pandemic in Ukraine	1247	75	1322
History of Ukraine	917	258	1175
Crimea	754	385	1139
Annexation of Crimea by the Russian Federation	335	167	502
2004 Ukrainian presidential election	367	58	425
Football in Ukraine	242	26	268
Bessarabia	204	51	255
2014 pro-Russian unrest in Ukraine	194	55	249
Communist Party of the Soviet Union	199	18	217
Christianity in Russia	142	30	172
History of Christianity in Ukraine	108	60	168
Flag of Ukraine	106	7	113
Alexander II of Russia	86	15	101
Eastern Front (World War II)	70	28	98
Bukovina	75	14	89
Epiphanius I of Ukraine	71	9	80
History of Crimea	48	14	62
Dissolution of the Soviet Union	52	4	56
Crimean Tatars	38	10	48
Catherine the Great	40	7	47
Culture of Ukraine	38	7	45
Abortion in Ukraine	41	3	44
Christmas in Ukraine	36	5	41
Armed Forces of Ukraine	32	3	35
Demographics of Ukraine	27	5	32
History of Kyiv	18	5	23
Eastern Front (World War I)	14	4	18
Foreign relations of Ukraine	15	3	18
Economy of Ukraine	13	2	15
Euromaidan	12	1	13
Galicia (Eastern Europe)	12	1	13
History of the Russian Orthodox Church	8	2	10
Government of Ukraine	6	1	7
Geography of Ukraine	6	0	6
Censuses in Ukraine	5	0	5
Government of the Ukrainian People s Republic in exile	3	0	3
Administrative divisions of Ukraine	3	0	3
Education in Ukraine	2	0	2
2022 Russian invasion of Ukraine	2	0	2
Buddhism in Ukraine	1	0	1
Total	5589	1333	6922

For the **large dataset :**

article	Not Weaponised Edits	Weaponised Edits	Total Edits
Eastern Front (World War II)	5651	1398	7049
Catherine the Great	5871	1035	6906
2014 pro-Russian unrest in Ukraine	3898	1275	5173
Annexation of Crimea by the Russian Federation	2763	1468	4231
Dissolution of the Soviet Union	3167	595	3762
Crimea	2508	1132	3640
Alexander II of Russia	2634	415	3049
Eastern Front (World War I)	2211	403	2614
Crimean Tatars	1937	533	2470
Economy of Ukraine	1951	338	2289
Armed Forces of Ukraine	1919	364	2283
Galicia (Eastern Europe)	1943	311	2254
History of Ukraine	1724	520	2244
Demographics of Ukraine	1771	230	2001
Euromaidan	1463	478	1941
Bukovina	1468	395	1863
COVID-19 pandemic in Ukraine	1588	110	1698
Bessarabia	1166	300	1466
Communist Party of the Soviet Union	1199	123	1322
Flag of Ukraine	1130	139	1269
History of Christianity in Ukraine	834	317	1151
Administrative divisions of Ukraine	833	190	1023
History of Crimea	726	254	980
History of Kyiv	696	233	929
Culture of Ukraine	786	116	902
2004 Ukrainian presidential election	660	89	749
Foreign relations of Ukraine	606	133	739
Government of Ukraine	624	64	688
Education in Ukraine	499	62	561
Christmas in Ukraine	433	90	523
Geography of Ukraine	455	49	504
History of the Russian Orthodox Church	363	125	488
Epiphanius I of Ukraine	307	45	352
Christianity in Russia	236	61	297
Football in Ukraine	242	27	269
Censuses in Ukraine	58	7	65
Buddhism in Ukraine	55	2	57
Abortion in Ukraine	44	3	47
Government of the Ukrainian People s Republic in exile	37	9	46
2022 Russian invasion of Ukraine	7	7	14
Total	56463	13445	69908

Table 5: Summary statistics extracted from extended the dataset.

References

[1] D. Fallis, "Toward an epistemology of wikipedia," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 10, pp. 1662–1674, 2008. DOI: https://doi.org/10.1002/asi.20870. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20870. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.20870.

[2] W. Foundation, *Wikimedia statistics — unique devices (all wikipedia projects)*, https://stats.wikimedia.org/#/all-wikipedia-projects/reading/unique-devices/normal|line|2-year|~total|monthly, Accessed: 2025-12-28, 2025.

[3] L. Kurek, C. Budak, and E. Gilbert, "Wikipedia in wartime: Experiences of wikipedians maintaining articles about the russia–ukraine war," in *Proceedings of the ACM on Human-Computer Interaction*, ser. CSCW, vol. 7, Association for Computing Machinery, 2023, pp. 1–27.

[4] D. Saez-Trumper, *Online disinformation and the role of wikipedia*, 2019. arXiv: 1910.12596 [cs.CY]. [Online]. Available: https://arxiv.org/abs/1910.12596.

[5] V. Makovska, "Vandalism or knowledge manipulation? detecting narratives in wikipedia edits," Master's thesis. Supervisors: Mykola Trokhymovych & Diego Saez-Trumper, M.S. thesis, Ukrainian Catholic University, Lviv, Ukraine, 2025. [Online]. Available: https://hdl.handle.net/20.500.14570/5697.

[6] M. Hidri, *Tracking weaponisation and culture heritage manipulation in ukraine by russia*, Available at: https://docs.google.com/spreadsheets/d/1vs_9Z2B-v2_QQBKnYhUz6LIKGePX7g2FtIABAvajpcc/edit?gid=0, 2025.