

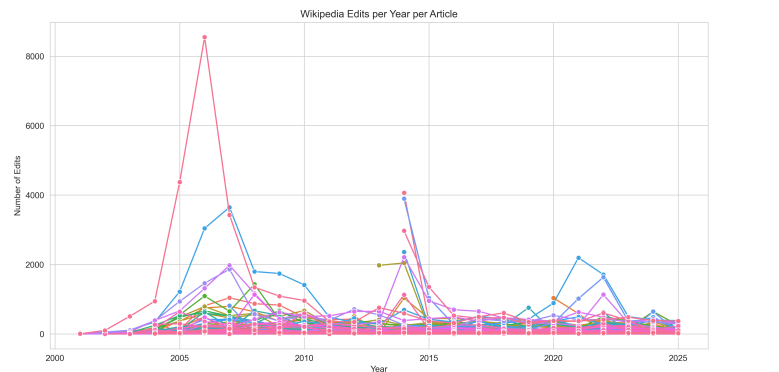


Results – Midterm Analysis (28 Oct. 2025)

1. Motivation

In the file `Wikipedia_articles.txt`, we gathered around **340 Wikipedia articles** related to **Ukrainian and Russian culture, history, and politics**.

By retrieving all the edits made to these articles, we aim to identify **patterns of activity**, for instance, peaks in editing activity during major historical or geopolitical events.



- 2004 Ukrainian presidential election
- 2014 pro-Russian unrest in Ukraine
- 2022 Russian invasion of Ukraine
- Abortion in Ukraine
- Administrative divisions of Ukraine
- Alexander II of Russia
- All-Ukrainian Union of Churches of Evangelical Christian Baptists
- Andrii Shevchenko
- Annexation of Crimea by the Russian Federation
- Architecture of Ukraine
- Armed Forces of Ukraine
- Armenian Diocese of Ukraine
- Art of Ukraine
- Arts of Ukraine
- Banking in Ukraine
- Baptism of Kyiv
- Baptists in Ukraine
- Belarusians in Ukraine
- Belarusians
- Bibliography of Ukrainian history
- Brotherhood of Independent Baptist Churches and Ministries of Ukraine
- Buddhism in Ukraine
- Bulgarians in Ukraine
- COVID-19 pandemic in Ukraine
- Carpathian Ruthenia
- Catherine the Great
- Catholic Church in Ukraine
- Cossacks in Ukraine
- Christianity in Russia
- Christianity in Ukraine
- Christmas Eve in Ukraine
- Christmas in Ukraine
- Crimea of Ukraine
- Coat of arms of Ukraine
- Communist Party of Ukraine (Soviet Union)
- Communist Party of the Soviet Union
- Constitution of Ukraine
- Corruption in Ukraine
- Cossack Hetmanate
- Crime in Ukraine
- Crimea
- Crimean Karaites
- Crimean Khanate
- Crimean Tatar language
- Crimean Tatar literature
- Crimean Tatars
- Crimian peninsula
- Culture of Russia
- Culture of Ukraine
- Declaration of Independence of Ukraine
- Demographics of Ukraine
- Direktorate of Ukraine
- Dissolution of the Soviet Union
- Dnipropetrovsk
- Dons (Crimea)
- Eastern Front (World War I)
- Eastern Front (World War II)
- Economic history of Ukraine
- Economy of Ukraine
- Education in Ukraine
- Electronics in Ukraine
- Electricity in Ukraine
- Encyclopedia of Ukraine
- Energy in Ukraine
- English Civil War
- Epiphany in Ukraine
- Ethnic groups in Ukraine
- Euroaidan
- Evangelical Lutheran Church in Russia, Ukraine, Kazakhstan and Central Asia
- Evangelical Presbyterian Church of Ukraine
- FC Dynamo Kyiv
- Flag of Ukraine
- Folklore of Ukraine
- Football in Ukraine
- Foreign relations of Ukraine
- Freedom of the press in Ukraine
- Galicia (Central Europe)
- Galicia (Eastern Europe)
- Gender inequality in Ukraine
- Geography of Ukraine
- Government of Ukraine
- Granting of autocephaly to the Orthodox Church of Ukraine
- HRWAD in Ukraine
- Health in Ukraine
- Hills of Kyiv
- Hilonging Church Kyiv
- Hilonging Ukraine
- Hindus in Ukraine
- Historical regions in present-day Ukraine
- History of Christianity in Ukraine
- History of Crimea
- History of Kyiv
- History of Ukraine
- History of UkraineEarly modern period
- History of UkraineWorld War II and the Nazi Occupation
- History of Ukrainian animation
- History of banking in Ukraine
- History of jewellery in Ukraine
- History of rail transport in Ukraine
- History of shipbuilding in Ukraine
- History of television in Ukraine
- History of the Cossacks
- History of the Jews in Ukraine
- History of the Russian Orthodox Church
- History of the Soviet Union (1985-1991)
- Holodomor
- Human rights in Ukraine
- Human trafficking in Ukraine
- Hungarians in Ukraine
- Humor in Ukraine
- Immigration to Ukraine
- Imperial Russian Army
- Intangible Cultural Heritage of Ukraine
- Invasion of Poland
- Islam in Ukraine
- Ivan Franko
- Jamaica
- Judiciary of Ukraine
- Khmelnitsky Uprising
- Khott Governance (Ukraine)
- Kiev
- Kiev Pechersk Lavra
- Kiev Voinodship
- Kievian Rus
- Kingdom of Galicia and Lodomeria
- Koloman Cossacks
- Kyiv
- Kyiv Pechersk Lavra
- LGBT rights in Ukraine
- Languages of Ukraine
- Latin Church in Ukraine
- Law of Ukraine
- Left-bank Ukraine
- Leonid Kuchuk
- Leonid Kuchma
- Lira (Ukrainian instrument)
- List of Intangible Cultural Heritage elements in Russia
- List of Intangible Cultural Heritage elements in Ukraine
- List of Metropolises and Patriarchs of KievMetropolitan of Kiev and All Ukraine
- List of Patriarchs and Metropolises of Ukraine
- List of UNESCO World Heritage Sites in Ukraine
- List of Ukrainian composers
- List of Ukrainian painters
- List of World Heritage Sites in Ukraine
- List of castles in Ukraine
- List of cities in Ukraine
- List of islands of Ukraine
- List of longest rivers of Ukraine
- List of mountains in Ukraine
- List of museums in Ukraine
- List of nature reserves of Ukraine
- List of places named after peopleUkraine
- List of waterfalls in Ukraine
- Literature of Ukraine
- Mass media in Ukraine
- Mass media of Ukraine
- Massacres of Poles in Volhynia and Eastern Galicia
- Media in Ukraine
- Media in Ukraine
- Ministry of Internal Affairs (Ukraine)
- Minorities in Ukraine
- Modern History of Ukraine
- Moldovans in Ukraine
- Museums in Ukraine
- Music of Ukraine
- Mykhailo Hrushevsky
- Mykola Kostomarov
- Mythology of Ukraine
- Name of Ukraine
- National Art Museum of Ukraine
- National Historical Museum of Ukraine
- National Museum of Folk Architecture and Folkways of Ukraine
- National Parliamentary Library of Ukraine
- National anthem of Ukraine
- National parks of Ukraine
- New Russia
- New Russia
- November Uprising
- Opera in Ukraine
- Opera in Ukraine
- Operation Revolution
- Organization of Ukrainian Nationalists
- Orthodox Church of Ukraine
- Outline of Ukraine
- Outline of the Russo-Ukrainian War
- Partitions of Poland
- Photography in Ukraine
- Podolia
- Poles in Ukraine
- Polish areas annexed by the Soviet Union
- Polish-Soviet War
- Political parties in Ukraine
- Politics of Ukraine
- Pop music in Ukraine
- Populated places in UkraineRural settlements
- Population transfers in the Soviet Union
- Post-Soviet transition in Ukraine
- President of Ukraine
- Principality of Kiev
- Prostitution in Ukraine
- Protestants in Ukraine
- Public holidays in Ukraine
- Racism in Ukraine
- Reichskommissariat Ukraine
- Religion in Ukraine
- Revolution of Dignity
- Right-bank Ukraine
- Roman Catholics in Ukraine
- Romanians in Ukraine
- Russia
- Russian Architecture
- Russian Civil War
- Russian Empire
- Russian Monorite
- Russian Orthodox
- Russian Orthodox Church
- Russian Orthodox Old-Rite Church
- Russian Revolution of 1917
- Russian invasions of Ukraine
- Russian literature
- Russians in Ukraine
- Russification of Ukraine
- Russo-Ukrainian War
- Russo-Ukrainian WarFull-scale Russian invasion of Ukraine (2022-present)
- Rusophelia
- Rusophone
- Saint Sophia Cathedral in Kyiv
- Saint Sophia Cathedral, Kyiv
- Science and technology in Ukraine
- Security Services of Ukraine
- Seven Natural Wonders of Ukraine
- Seven Wonders of Ukraine
- Seven Wonders of UkraineSeven Natural Wonders
- Steboda Ukraine
- Soviet Top League
- Soviet Ukraine
- Soviet Union
- Soviet peasent
- Sport in Ukraine
- St. Nicholas Roman Catholic Cathedral, Kyiv
- Stock exchanges in Ukraine
- Sub-Carpathian Reformed Church
- Sudbina in Ukraine
- Sviatoslav I of Kiev
- Symbols of Ukraine
- Taras Shevchenko
- Telecommunications in Ukraine
- Television in Ukraine
- The Church of Jesus Christ of Latter-day Saints in Ukraine
- The Ukrainian Week
- Theater in Ukraine
- Tourism in Ukraine
- Transport in Ukraine
- Turks in Ukraine
- USSR
- USSR national football team
- Ukraine
- Ukraine after Russian Revolution
- Ukraine after the Russian Revolution
- Ukraine during World War I
- Ukraine national football team
- Ukrainian Bible Society
- Ukrainian Greek Catholic Church
- Ukrainian Greeks
- Ukrainian Insurgent Army
- Ukrainian Lutheran Church
- Ukrainian Museum
- Ukrainian National Chernobyl Museum
- Ukrainian National Committee
- Ukrainian Orthodox Church (Moscow Patriarchate)
- Ukrainian Orthodox Church - Kyiv Patriarchate
- Ukrainian Orthodox Greek Catholic Church
- Ukrainian Premier League
- Ukrainian Rusephiles
- Ukrainian SSR
- Ukrainian Soviet Socialist Republic
- Ukrainian State
- Ukrainian War of Independence
- Ukrainian animation
- Ukrainian architecture
- Ukrainian art
- Ukrainian avant-garde
- Ukrainian baroque
- Ukrainian cuisine
- Ukrainian culture
- Ukrainian dance
- Ukrainian diaspora
- Ukrainian embroidery
- Ukrainian fairy tale
- Ukrainian folk music
- Ukrainian folklore
- Ukrainian historical regions
- Ukrainian hymnia
- Ukrainian language
- Ukrainian literature
- Ukrainian national clothing
- Ukrainian people
- Ukrainian underground
- Ukrainian wedding traditions
- Ukrainian wine
- Ukrainian wreath
- Ukrainians
- Ukrainians in Estonia
- Unification council of the Orthodox churches of Ukraine
- Vernadsky National Library of Ukraine
- Viktor Yanukovich
- Viktor Yushchenko
- Volhynia
- Volhynia genocide
- War in Donbas
- Western Ukrainian Cergy
- Wildlife of Ukraine
- Women in Ukraine
- World Network of Biosphere Reserves in Europe and North AmericaUkraine
- World War II
- Zaporozhian Cossack
- Zaporozhian Cossacks

Observed Peaks of Activity

Distinct peaks can be identified around several key periods:

- 2005–2008** → Post–Orange Revolution period, marked by Russo-Ukrainian gas disputes and general geopolitical tension. This era also coincides with Wikipedia's "golden age," with roughly **50,000 active users** at its 2007 peak.
- 2014** → The **Russian annexation of Crimea**.

- **2020–2021** → The **COVID-19 pandemic**, which likely increased online activity overall.
- **2022–2025** → The **large-scale Russian invasion of Ukraine**, generating global attention and narrative conflict.

These observations motivate a deeper investigation into **editor behavior and content modification patterns**.

2. Research Structure

This midterm analysis is divided into two complementary parts:

A. Weaponised vs. Non-Weaponised User Analysis

A global overview of editing patterns among a sample of **2,454 unique users**, distinguishing between edits classified as *weaponising* or *non-weaponising*.

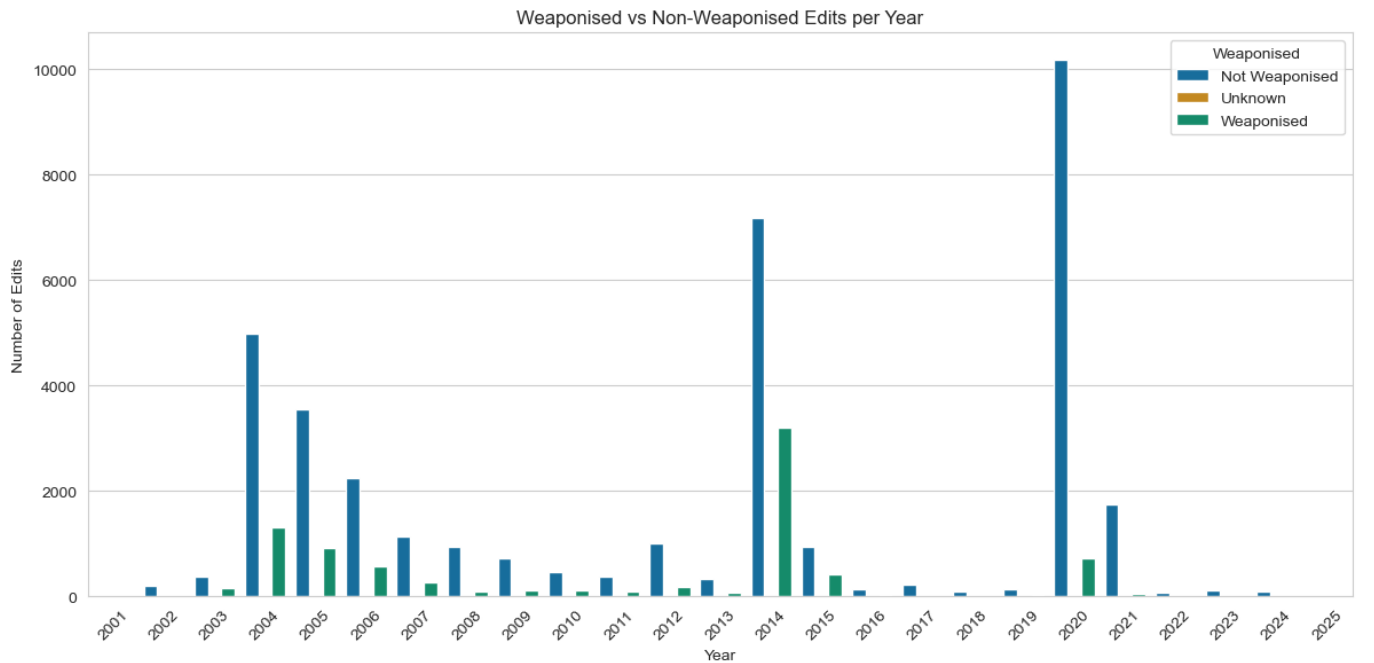
B. Weaponising Users — Fine-Grained Analysis

A focused investigation of **400 unique users** from the *fine-grained dataset*, restricted to edits identified as weaponising by the language model.

3. Weaponised vs. Non-Weaponised User Analysis

This section is based on the file `matched_edits_all.csv`, produced by **Emanuela**, which aligns fine-grained edits with corresponding user data from the *LLM Results* folder.

One can first check how *Weaponising* VS *Non weaponising* edits are distributed over time.



There is a clear correlation with the first graph, showing more edits on crucial geopolitical periods as well as a peak of weaponsing edits. No peaks are displayed for the 2022–2025 periods since no article in the dataset was about Russian large scale invasion in Ukraine.

Metadata of the `matched_edits_all.csv`

```
{'total_edits': 45697,  
 'total_unique_users': 2454,  
 'number_unique_articles': 41,  
 'num_bot_users': 92,  
 'num_anonymous_users': 1011,  
 'num_registered_users': 1351,  
 'num_bot_edits': 1376,  
 'num_anonymous_edits': 8315,  
 'num_registered_edits': 36006,  
 'weaponised_edits': 8388,  
 'non_weaponised_edits': 37278,  
 'unknown_weaponised_edits': 31}
```

There is a lot we can analyse coming from this metadata. We first start to focus on the **registered** users and specially the top 10 most frequent editors among the 1351 registered users.

Top 10 users general info

The top 10 registered editors are :

```
['LSGH',  
 'Toddy1',  
 'TheDraperyFalls',  
 'UP9',  
 'Genyo',  
 'Aleksandr Grigoryev',  
 'Dbachmann',  
 'Vasile',  
 'Irpen',  
 'Mzajac']
```

Proportion regarding Ur/Ru articles

To identify how much each of the top 10 users focused on Ukraine- or Russia-related content, I filtered their edited articles using both small and large sets of keywords related to these two countries.

```
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",
```

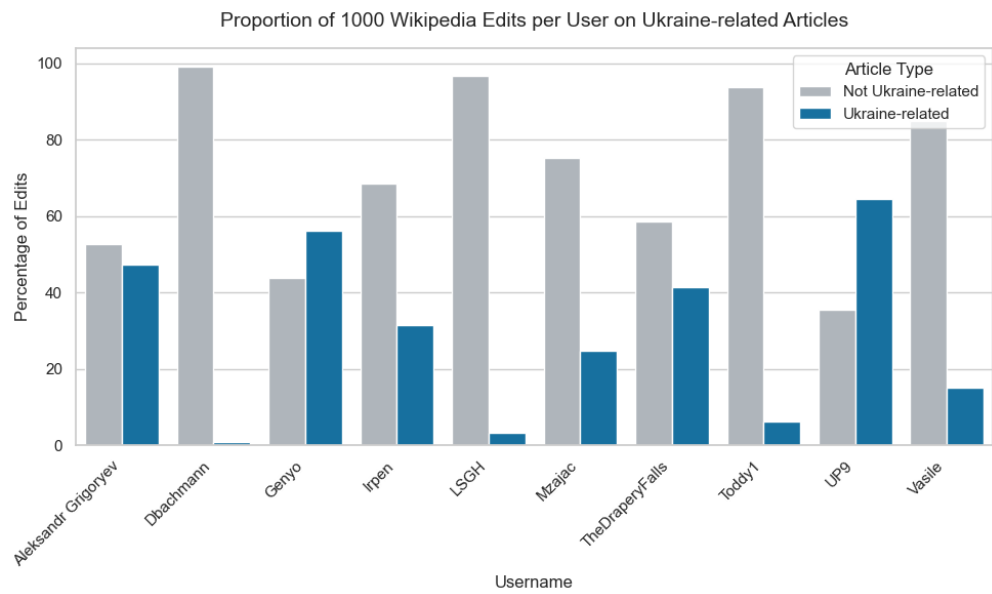
```
"donbas", "donetsk", "luhansk",  
    "maidan", "yanukovych", "yushchenko", "zelenskyy", "poroshenko",  
"catherine", "bukovina", "bessarabia",  
    "eastern", "euromaidan", "dnipro", "odessa", "sevastopol", "putin",  
"rus'", "russia", "russian", "moscow",  
    "kremlin", "soviet",  
  
# politics & 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", "luts'k", "uzhhorod",  
"dnipropetrovsk", "donetsk oblast",  
    "luhansk oblast", "transcarpathia", "prikarpatia", "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",

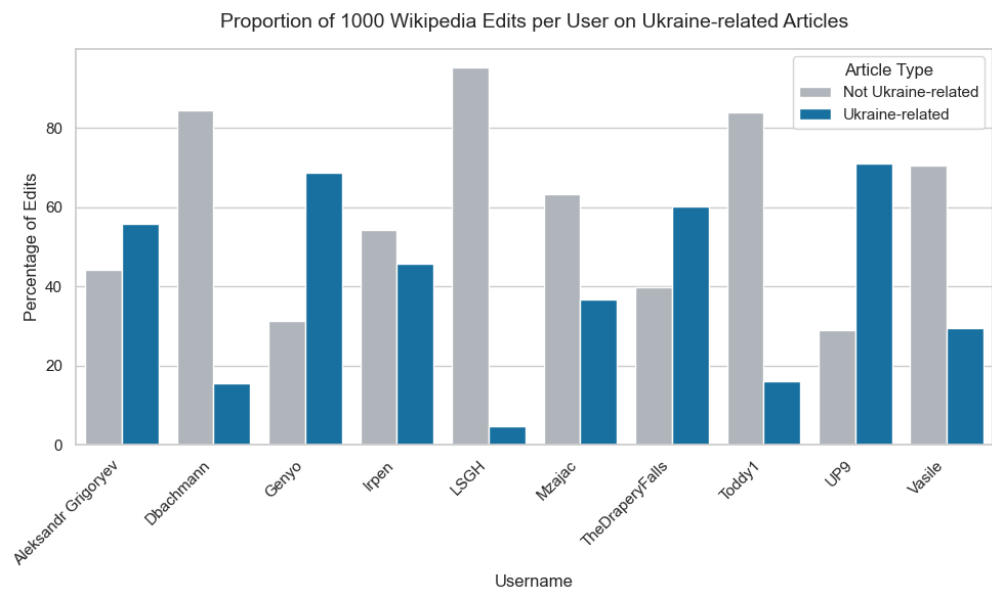
# other
"gas pipeline", "north stream", "energy crisis", "grain corridor",
"black sea fleet", "peace talks",
"donetsk people's republic", "luhansk people's republic", "kyiv
oblast", "liberation", "resistance",
"occupation zone", "ukrainian refugees", "mariupol steel plant",
"azovstal", "bucha", "irpin", "kharkiv offensive"
]
```

I did the analysis for a sub sample of 1000 edits per user and for all the edits the user made since the creation of his account using Wikipedia's user API.

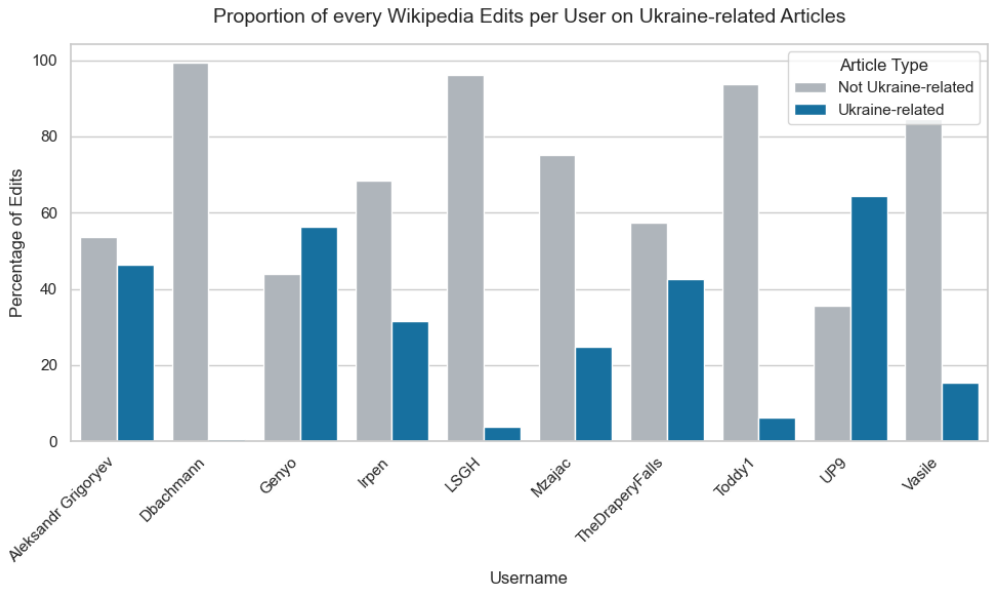
With the small set for 1000 edits per user :



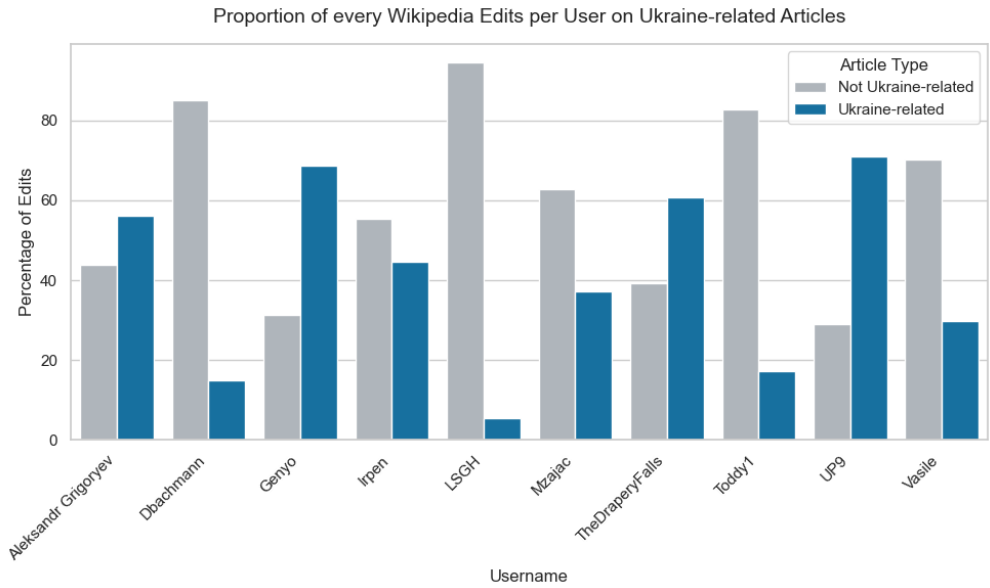
With the large set for 1000 edits per user :



With the small set for every edits per user :



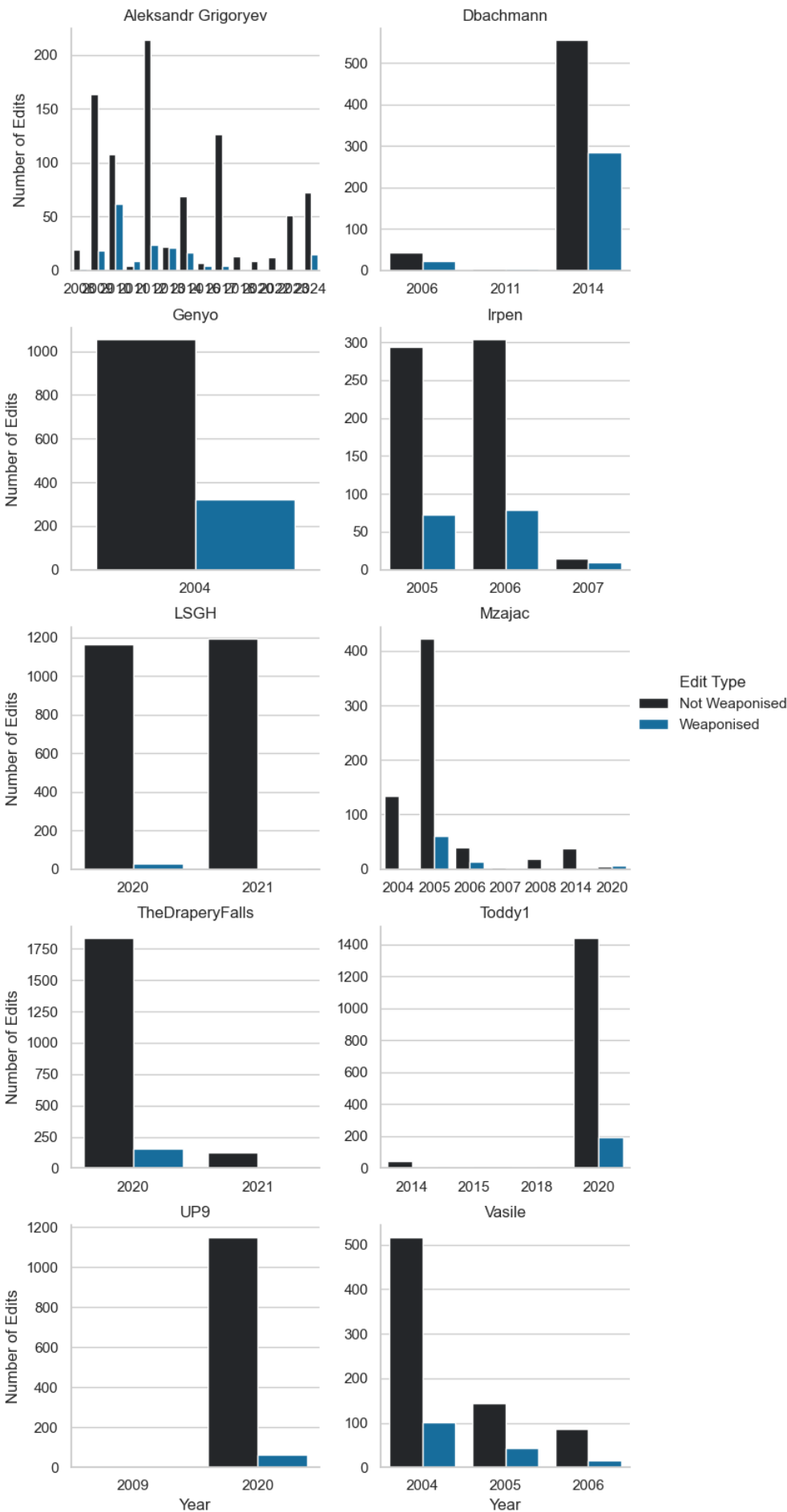
With the large set for every edits per user :



We can see that while the batche's size of edits doesn't really make any difference, the set's size of Ua/Ru keywords does. The small size really focus on Ua/Ru specific article while the large set is more about controversial/political/historical article in general. Now, we can select only the Ua/Ru related articles to see if a certain pattern shows off.

Weaponising edits vs non weaponising edits per year for top 10 for only Ur/Ru related article

Weaponised vs Non-Weaponised Edits per Year — Top 10 Users



The graph shows that while Dbachmann seldom edits Ukraine- or Russia-related pages, those few edits are more likely to be weaponising compared to others. In general, these top users are broad Wikipedia contributors, with limited focus on Ukraine/Russia topics and little evidence of weaponising behaviour.

Groups

| Group | Who it includes | Typical criteria | Can edit semi-protected pages? | Can edit extended-protected pages? |
|-------------------|-------------------|----------------------|--------------------------------|------------------------------------|
| * | Anyone (even IPs) | none | ✗ | ✗ |
| user | Registered users | account created | ✓ | ✗ |
| autoconfirmed | Older + ≥10 edits | ~4 days, ≥10 edits | ✓ | ✗ |
| extendedconfirmed | Sustained editors | ≥30 days, ≥500 edits | ✓ | ✓ |

Note: hard to get some relevant information for only the top 10 users. Maybe will me more relevant when focus only on the weaponising users for the fiengrained file.

Registered vs IP adress users

The file, as displayed in the metadata, includes 2,454 unique users, of which:

- 1,351 are registered accounts
- 1,011 are anonymous (IP addresses)
- 92 are bots

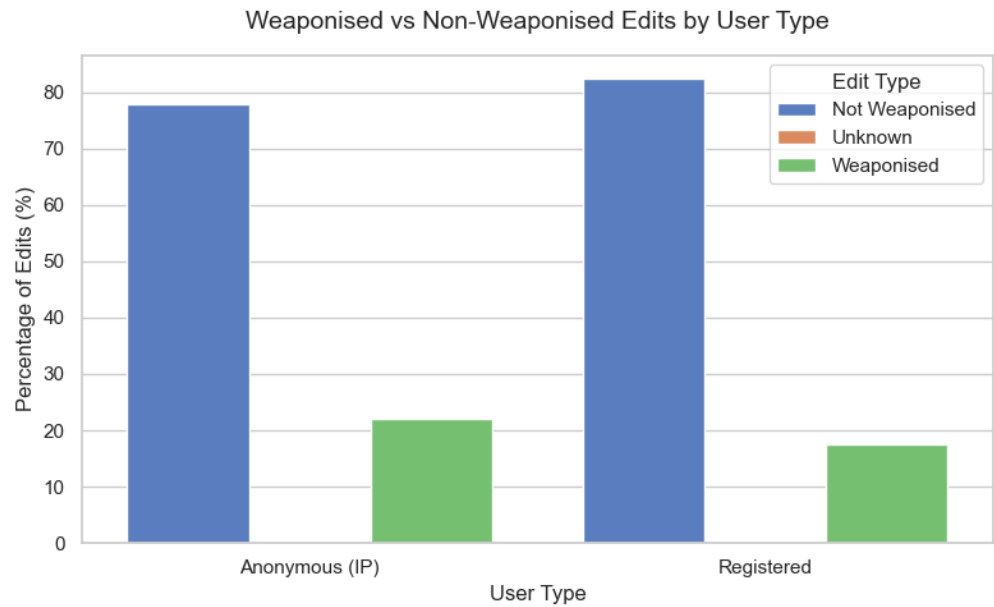
Although the number of registered and anonymous users is roughly comparable, the volume of edits differs significantly:

- Registered users: 36,006 edits
- Anonymous users: 8,315 edits

This implies that registered users contribute over four times more edits than anonymous users. Thus we can raise the following research question

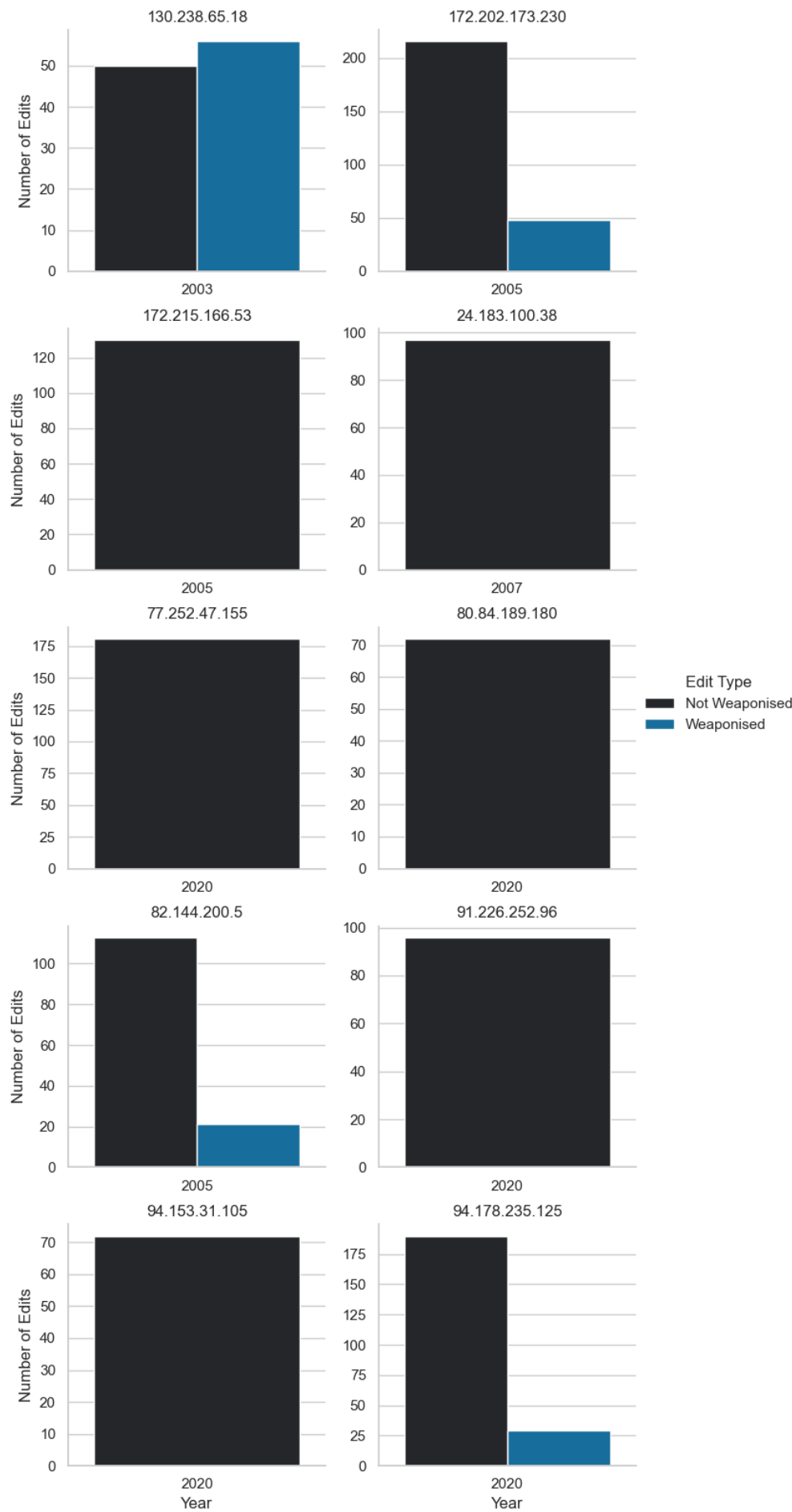
Are anonymous (IP address) users more likely to produce weaponising edits compared to registered users due to their anonymity?

The graph below shows that in proportion, IP adress user are slightly more weaponsing than registered user. They feel maybe more confortable to weaponise under anonymity.



As we did for the top 10 registered users, we can also see how much edits and when the top 10 anonymous users are weaponising or not. As seen in the plot, we can see that the numbers of edits is way lower than the top 10 registered user and all of the top 10 IP addresses without exception are only editing during a **single** year. As seen previously, registered users are the most of them editing for a longer time frame.

Weaponised vs Non-Weaponised Edits per Year — Top 10 Anonymous Users



By using an IP Geolocation API, we can check whether the used IP address is from a proxy or not and where the edit is located :

 [View Interactive IP Map](#)

4. Weaponised User Analysis

After analysing weaponised vs. non-weaponised edits for both registered and anonymous users, we can now go a step deeper and focus exclusively on the **users who have been classified as weaponising**. This allows us to explore not only *who* is weaponising, but also *how* different types of users (registered vs. anonymous) engage in specific forms of weaponisation.

Note: Unfortunately, I was not able to retrieve either the articles or the timestamps corresponding to the edits.

Metadata of the `matched_edits_all.csv`

```
{'total_edits': 513,  
 'total_unique_users': 397,  
 'num_bot_users': 9,  
 'num_anonymous_users': 146,  
 'num_registered_users': 242,  
 'num_bot_edits': 11,  
 'num_anonymous_edits': 156,  
 'num_registered_edits': 346}
```

From the `fg_user_known.csv` file, we identify almost 400 weaponising users — including 146 anonymous (IP-based) editors and 242 registered editors.

Top 10 users general info

The top 10 registered editors are :

```
['Aleksandr Grigoryev',  
 'RGloucester',  
 'Yulia Romero',  
 'Dbachmann',  
 'Lvivske',  
 'JASpencer',  
 'Kuban kazak',  
 'Irpen',  
 'Mzajac',  
 'Seryo93']
```

As expected, no IP addresses appear in this top 10. Some usernames have already appeared in the previous ranking, suggesting that a few users are consistently active across different subsets of edits. One can may prove those kind of behaviour in the next chapter "Graph and Network Visualisation".

To further characterise editing behaviour, we can ask:

Do IP address editors are more concise and harmful than registered users ?

This global RQ can be answered and sub-divided in 4 themes :

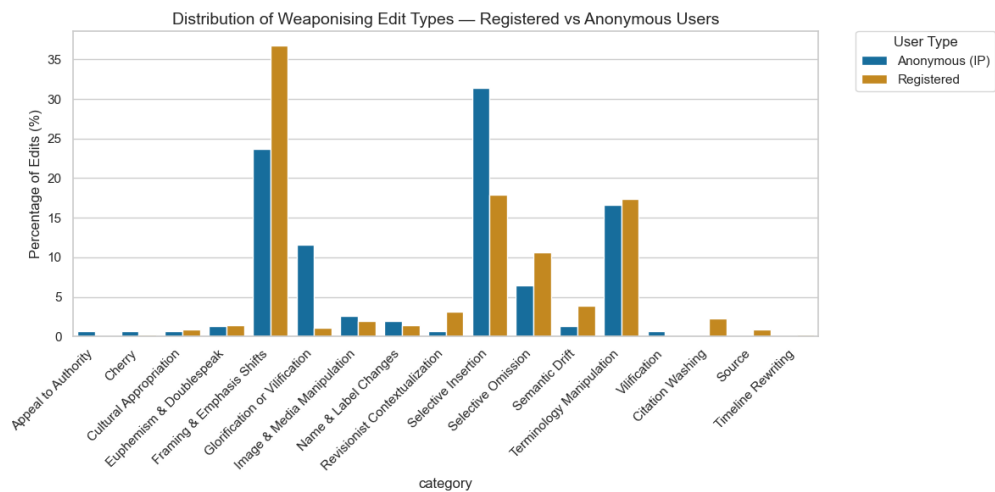
- A. the category of weaponisation by user type
- B. the edit magnitude
- C. the direction of the edit (addition or deletion)
- D. The semantic and lexical analysis

A. Category of weaponisation per user type

This category is driven by the following RQ:

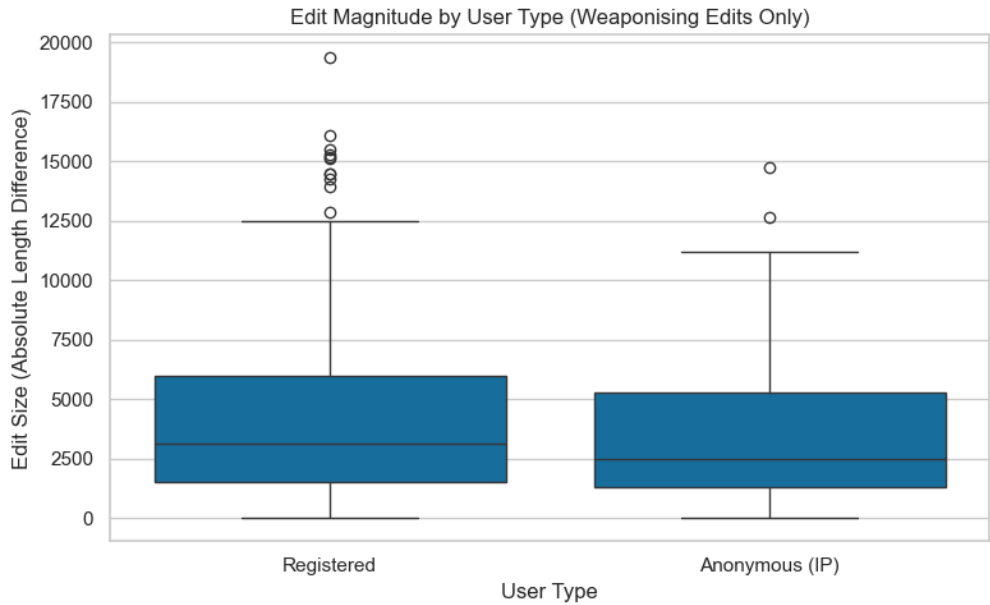
How do registered and anonymous users differ in the type of weaponising strategies they employ?

The graph below shows an interesting distribution of weaponisation categories across user types. On one hand, registered users appear to favour **Framing and Emphasis Shifts*** : a technique that often involve adding context, rewording, or subtly reframing information to promote a particular point of view. These users generally produce longer edits and seem to embed manipulative intent in more sophisticated ways. On the other hand, anonymous users (IPs) tend to rely more on **Selective Insertion**, a strategy that typically consists of short, targeted, and often more direct interventions. This pattern suggests that anonymous users might engage in quicker, less nuanced, but potentially more disruptive forms of manipulation.



B. Edit Magnitude

In order to emphasize more this behaviour, one can perform a boxplot of edit magnitude per user type by checking the length of the `changed_version` vs the length of the `initial_version` :



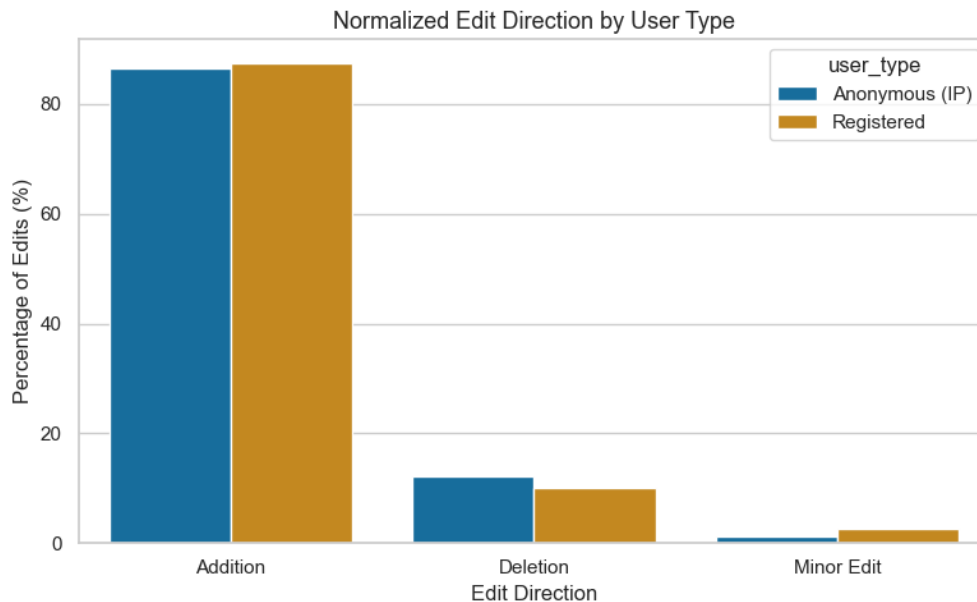
The median weaponising edit size was comparable between registered and anonymous users, though registered users exhibited a greater variance and more extreme outliers. This suggests that while both groups engage in manipulative micro-edits (e.g., terminology shifts or short insertions), registered users occasionally perform larger structural interventions — possibly reframing or rewriting entire paragraphs. In contrast, anonymous users tend to contribute smaller, localized edits (consistent with quick, opportunistic interventions).

Maybe a good idea to assess the statistical significance of this results by calculating the p-value for both groups. So we suppose that the two groups does not provide a significal difference. By using Mann-Whitney U test i.e `mannwhitneyu` from `scipy.stats` module, we find a p-value of **0.03180**. This indicates that the observed difference between the two groups would occur by chance less than 3% of the time — providing moderate evidence of a statistically meaningful distinction in their editing behaviours.

C. Change Direction

We can classify edits as Additions (positive Δlen), Deletions (negative Δlen), or Rewrites (\pm small Δlen) and compare distribution by user type. By folowing this question :

Are IPs more likely to add new sentences (insertions) or remove content (deletions)?



D. Lexical & Semantic Analysis

Possible to analyse the sentiment polarity and toxicity by using model like [cardiffnlp/twitter-roberta-base-sentiment](#).

ANNEX : Track the .csv files

```
all_user_profiles.csv           # Keeps the top10 user metadata gathered
with WP user API.
matched_edits_all.csv          # All the ngrams that matched with a
weaponising or non weaponising edits.
all_user_contribs_sub_n=1000.csv # Sub-sample of 1000 contributions of
top10 user gathered by WP user contrib API.
all_user_contribs.csv           # All contributions of top10 user
gathered by WP user contrib API.
df_user_known.csv               # All weap username matched in the
finegrained2
df_user_unknown.csv             # All the weap username not found in the
finegrained2
fg_user_known_delux.csv         # ll weap username matched in the
finegrained2 with extra info for analysis
```