From Fringe to Frontline

/- Telegram Trends in Austria's 2024 Elections

Background and Key Findings

The 29 September 2024 National Council elections in Austria were a pivotal moment in the country's political history as, for the first time, the populist far right Freedom Party (FPÖ) became the biggest party in Parliament; A total of 28.8 per cent of the valid votes cast resulted in 57 (31.1 per cent) of the seats in the lower house, the 183-member National Council. The two traditional political parties that have dominated Austrian post-war politics – the Austrian Peoples Party (ÖVP) and the Social Democratic Party (SPÖ) – have been facing an increasingly fragmented political environment. Overall, the FPÖ has been steadily gaining support through its anti-immigrant, anti-EU, and anti-establishment platform, and it appears poised to gain further momentum. Those coalition talks since collapsed and the FPÖ leader Kickl has since been assigned by the President to lead coalition talks (with the ÖVP).

The 2024 electoral campaign was assessed as fair without any foreign or party orchestrated disinformation campaign detected.[1] AUF1, labelled by Austrian domestic intelligence as an "alternative right-wing extremist" media channel, alleged electoral fraud risks, claiming postal voting could prevent an FPÖ victory.

---

[1] Austrian and German media did not report of any orchestrated disinformation campaigns in the Austrian elections. See: https://www.sueddeutsche.de/politik/oesterreich-nationalratswahl-2024-fpoe-oevp-spoe-rechtsruck-koalition-regierungsbildung-lux.XLzNMBzLUQnnp3p7MSmnPs; https://www.bpb.de/kurz-knapp/hintergrund-aktuell/552357/nationalratswahl-in-oesterreich-2024/; fact checking organisations only detected minor issues: https://gadmo.eu/vor-nationalratswahl-in-sterreich-falschbehauptungen-ber-bilanz-von-kanzler-nehammer-im-umlauf/

It further suggested a "deep state" plot to steal a win from the FPÖ. After election day, FPÖ leader Herbert Kickl cited "election manipulation" in his speech, expressing frustration over his party's exclusion from coalition talks led by the president and other parties.

Social media platforms and messaging services have become a crucial battleground in Austrian political campaigns, [mirroring trends seen globally](#). Parties and candidates have been leveraging [platforms](#) such as Facebook, Instagram, TikTok, and X to reach voters, particularly younger ones, who primarily consume news online and engage with politics via social media. One of the messaging services that has started playing an increasingly important role in Austria is Telegram, which provides parties like the FPÖ and right-wing groups a platform to directly communicate with supporters without facing the content restrictions of mainstream social media. For supporters sceptical of traditional media, Telegram has become a source for alternative perspectives, frequently amplifying narratives around topics like immigration, EU policies, and electoral integrity. In this study, we examine toxicity, hate speech, and extremism in Austrian Telegram groups ahead of the election.

Key Findings

/ **Key Channels:** The channels Eva Herrman Offiziel, Oliver Janich, and Uncut News were among the most active and influential, contributing significantly to the observed toxic, hateful, and, in some cases, extremist discourse. These channels are focal points in the network of far-right activists and may have driven a substantial portion of the concerning content.

/ **Prevalence Patterns:** Toxicity is the most common type of problematic content on the channels, followed by hate speech, and then extremism, which appeared less frequently. This suggests that the use of inflammatory and offensive language was widespread, though extreme ideological content was comparatively rare.

/ **Channel-Specific Trends:** Certain channels, like *auf1TV*, had high toxicity levels, but ranked lower for hate speech or extremism. This indicates that some channels are vehicles for only certain kinds of problematic content (see below for the difference between hate speech and toxicity, for example), and warrant tailored analysis of the themes and tone of discourse, indicating differences in content style and focus across channels.

/ **Influence of Specific Examples:** Posts with high hate speech, toxicity, and extremism probabilities (e.g., those from Martin Sellner, an Austrian far right extremist with the Telegram channel called *martinsellnerIB*)[2] exemplify how certain channels emphasised themes aligned with extreme or hostile ideologies, making them key points of interest for understanding the spread and impact of such discourse.

Methodology

The data for this project was collected and analysed by Austrian security researchers and wahlbeobachtung.org. The researchers collected and analysed data from a list of 27 Telegram channels provided by wahlbeobachtung.org. The Telegram channels were selected based on existing research and information available from Austrian state services (the **Directorate of State Security and Intelligence** and the **Federal Office for Cult Affairs**) and civil society organisations, such as the Documentation Centre of Austrian Resistance (**DÖW**), on media and social media covering and distributing disinformation and hate speech, including via Telegram channels.[3] Furthermore, the channels were selected because the

---

[2] Martin Sellner is the former leader of the Identitarian Movement Austria, who received donations from the Christ Church mass shooter. Since 2021, the display of symbols and gestures of the Identitarian Movement are prohibited in Austria.

[3] Bundesstelle für Sektenfragen, "**Ende der Maßnahmen – Ende des Protests? Das Telegram-Netzwerk der österreichischen COVID-19-Protestbewegung und die Verbreitung von Verschwörungstheorien**", April 2024.

above-mentioned reports concluded that these were the most influential in Austrian political public discourse and, therefore, are most relevant for the present analysis.

Data was collected from the Telegram Application Programming Interface (API), and covered the period from 1 September to 15 October 2024. The analysis focussed on detecting extremism, hate speech, and toxicity, using machine learning classifiers developed by the researchers. To get an overview of all the content being distributed on the selected channels, both posts by the owner of the respective channel and comments by other users were included in the analysis, resulting in a total of 9,313 observations (posts and comments) included in the analysis.

While "toxicity" and "hate speech" are closely related, they are not interchangeable, and can even occur independently of each other. To distinguish between the two categories, the following definitions underlie the automated detection models:

/ **Toxicity** indicates the potential of a comment to encourage aggressive responses or trigger other participants to leave the conversation.

/ **Hate speech** is defined as any form of expression that attacks or disparages persons or groups by characteristics attributed to the groups.

/ **Extremism** is any form of extreme or radical (political or religious) statement, or a fringe attitude or aspiration.

The researchers developed Toxicity, Hate Speech, and Extremism detection tools were used to analyse the social media content. The basis of the detection tools is

XLM-RoBERTa[4], a high-performing language model that is trained on multilingual data. The experts further pre-trained this model with additional unlabelled data to better capture current social media slang and phrasing. For the target tasks of hate speech and toxicity detection, the model was fine-tuned with human-annotated data.

Findings

Below, we show the prevalence of hate speech, toxicity, and extremism on the different channels. To illustrate this further, we give examples of posts (translated) that had particularly high scores on the respective scales. The analysis shows that there are significant differences both across Telegram channels and across the prevalence of the potentially concerning content. Toxicity appears to be the most prevalent, followed by hate speech. Based on the results of the deployed models, extremism is much less prevalent. This does not, however, mean that extremism on Telegram is not a problem or that we should not be concerned about the distribution of extremism via Telegram.

4 Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Giullaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, Veselin Stoyanov. 2020. "**Unsupervised Cross-lingual Representation Learning at Scale**", arXiv, Cornell University, 8 April 2020.

Graph 1: Overview of all analysed Telegram channels and the volume of posts
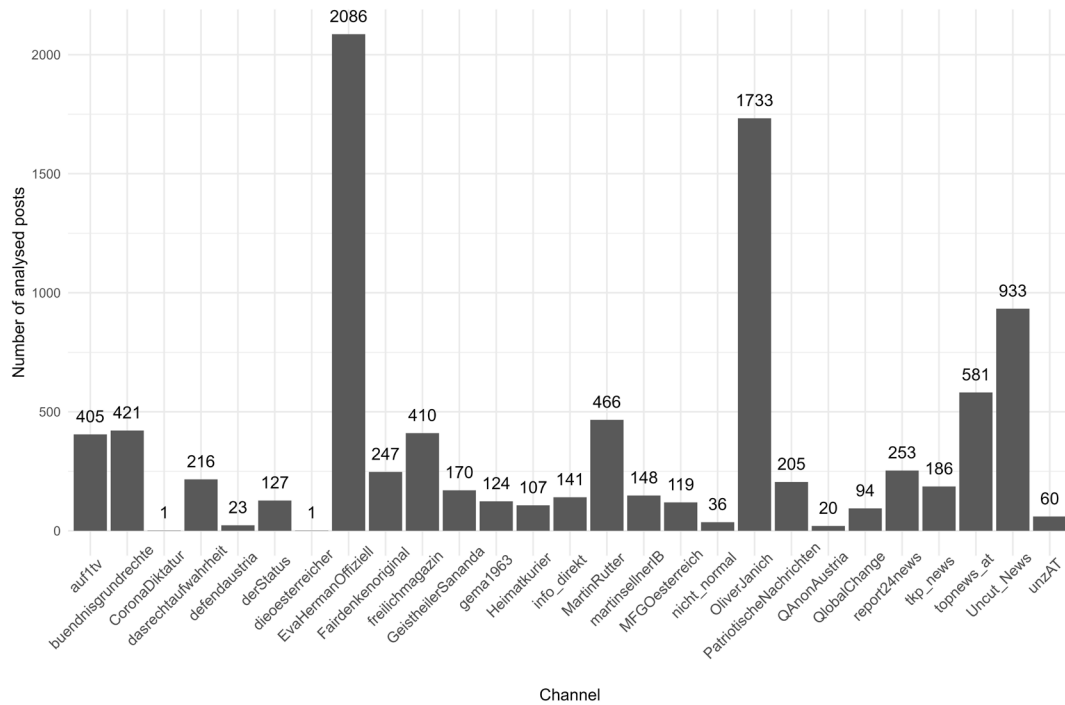

Graph 1 provides an overview of all analysed Telegram channels and the volume of posts included from each channel. The three channels with the highest number of posts — **Eva Herrman Offiziel, Oliver Janich,** and **Uncut_News** — stand out as particularly active compared to the others. Their elevated post counts, together with the high rate of views per post, suggest they were focal points for communication and likely influential in the discourse being studied. Other channels that could warrant further examination, based on high post-view rates, include **auf1TV, QlobalChange,** and **MartinSellnerIB**. This high activity could indicate that these channels are key sources of content around the topics of hate speech, extremism, and toxicity, and thus merit closer attention in gaining a better understanding the trends or patterns in this space. A deeper examination in our upcoming full report will explore the nature of the interactions on each channel, also taking into account more active engagement, such as commenting on posts. On the other hand, **CoronaDiktatur** and **dieoesterreicher** had only one post each

during the same period. Due to their limited data, these channels are excluded from the subsequent analyses.

Graph 2: Average views per post
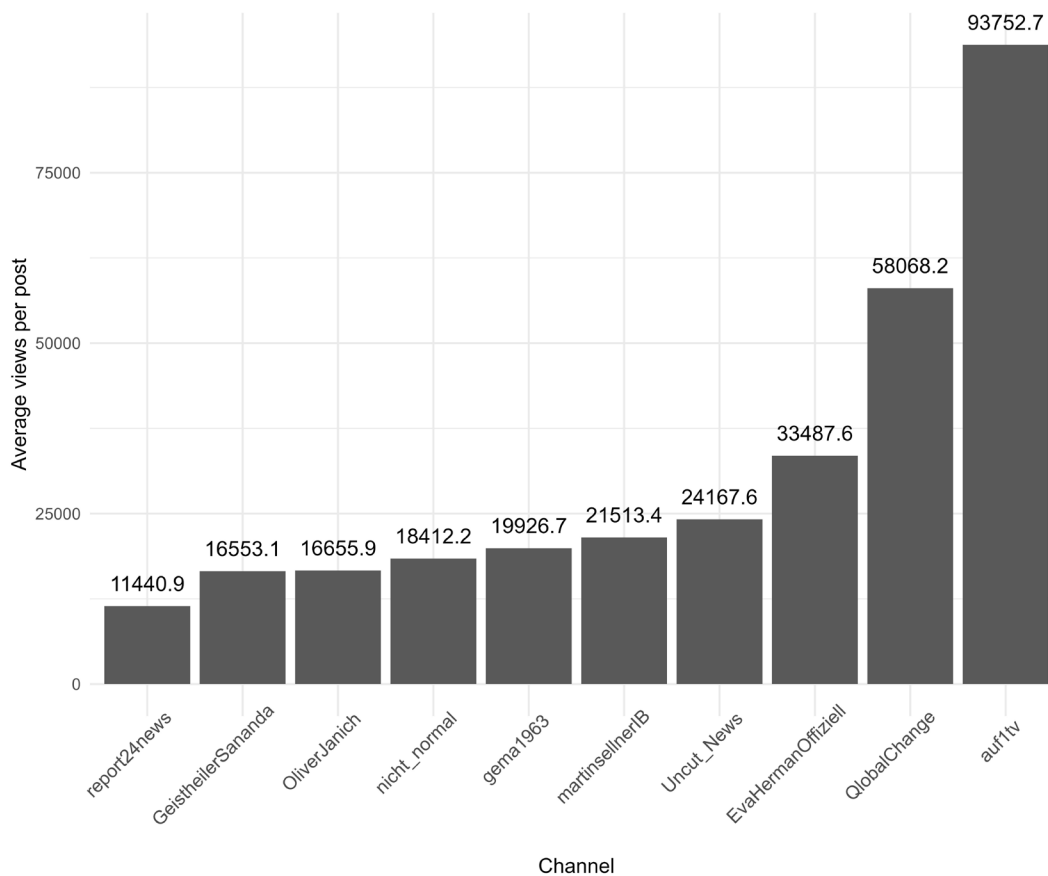
Graph 2 shows the average views per post for the ten channels with the highest average views. The views per post provide important insight into the reach and potential impact that these channels had. The more people view a post, the more likely it is that the content of the post will also spread beyond the Telegram channel it is originally from. A high number of views and engagement is nothing worrying on its own, but is problematic if combined with a high average probability of containing problematic posts, which increases the likelihood of the negative im-

pact of this content within and beyond Telegram.[8] The channels **auf1tv, Qlobal-Change,** and **EvaHermmanOffiziell** had the highest number of average views in our sample. Neither **QlobalChange** nor **EvaHermmanOffiziell** reappear on the channels with the highest probabilities of hate speech, toxicity, or extremism. This is a positive sign, as it suggests that the channels with the most engagement are not necessarily the channels with the most problematic content. **Auf1tv** does reappear as the channel with the highest average probability of toxicity. This indicates that the channel spread toxic content within and, quite likely, beyond Telegram.

Graph 3: Illustration of the distribution and likelihood of hate speech content
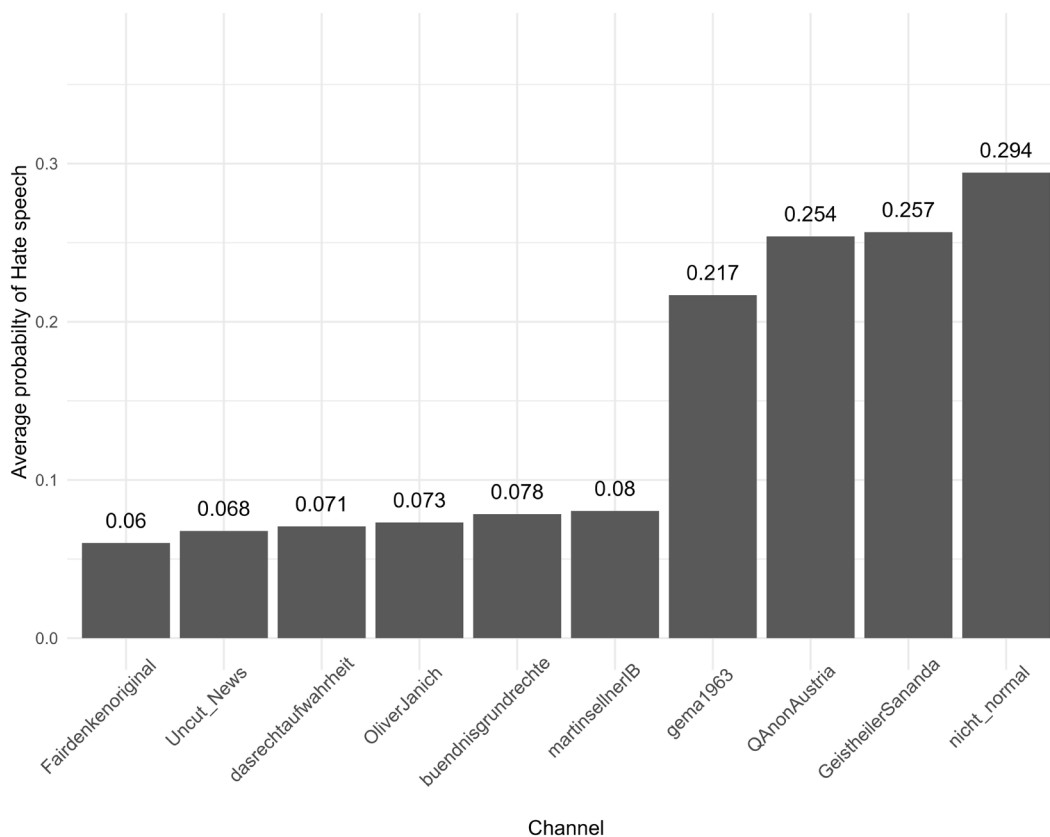
8 (Herasimenka, et al. 2022) Aliaksandr Herasimenka, Jonathan Bright, Aleksi Knuutila, & Philip N. Howard, "Misinformation and Professional News on Largely Unmoderated Platforms: The Case of Telegram", Journal of Information Technology & Politics, vol. 20, no. 2, 25 May 2022, pp. 198–212.

Graph 3 provides an illustration of the distribution and likelihood of hate speech content on some of the observed Telegram channels, focussing on the top ten. Channels with consistently higher probabilities suggest a sustained tendency towards hate speech, making them critical points for understanding the spread and impact of hate speech within this network. It is interesting to note that two of the most prolific channels (in terms of posting), **Oliver Janich**, and **Uncut_News**, were also among the top 10 for likelihood of hateful content. This finding suggests a need for further investigation into those channels, and specifically into the types of hateful content that they disseminate, and the targets thereof.

Graph 4: Average probability that a post includes toxicity for the ten channels with the highest average probability
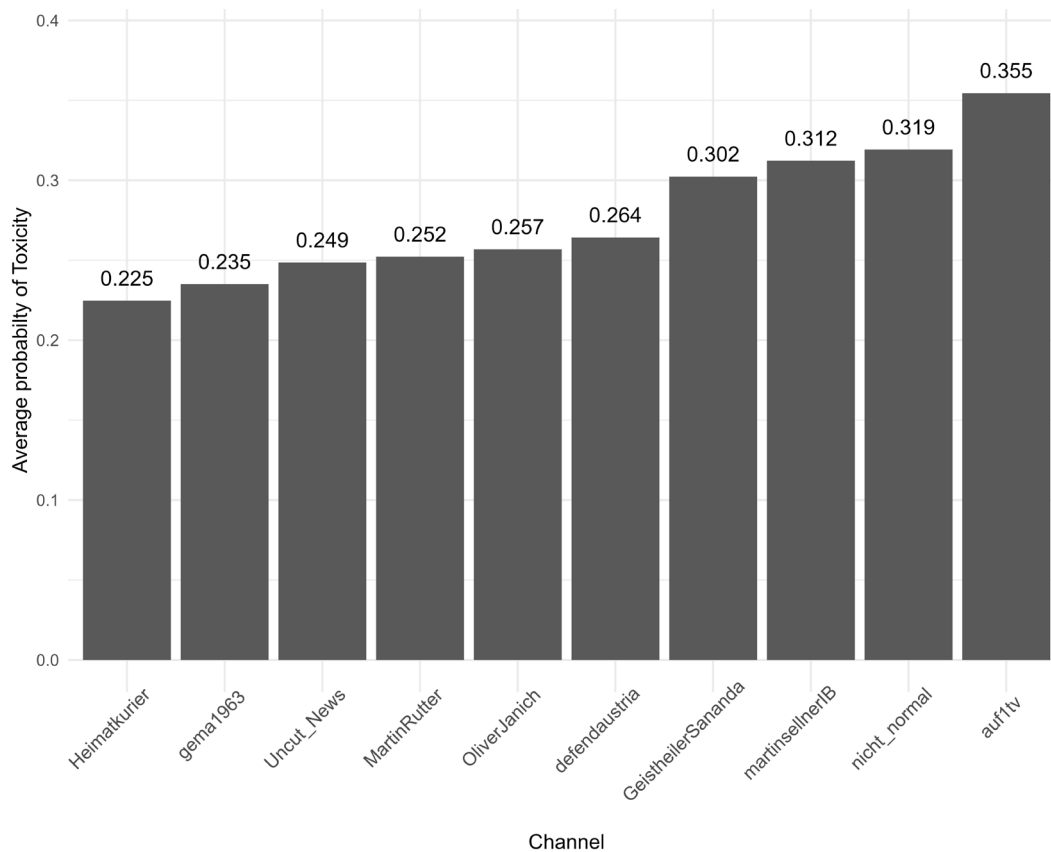
Graph 4 shows the average probability that a post included toxicity for the ten channels with the highest average probability. Of particular note here is the presence of **auf1TV** with the highest average likelihood of toxic content in a post, since **auf1TV** is not present at all in the top 10 for hate speech or extremist content. A deeper investigation could look into the nature of the posts on this channel, the topics that trigger the toxicity, and what makes them toxic but not hateful. An example of a post scoring high on the toxicity detector is the following one, from the channel **buendnisgrundrechte**. In this post, a conservative politician is called a slut and accused of lying.

Example 1: Toxic post from channel buendnisgrundrechte

'This ÖVP political slut blatantly lies to a voter to her face. She just mouths off at a factually asked question. 😡😡😡'

> [3] "Diese ÖVP-Politschlampe lügt einer Wählerin unverhohlen mitten ins Gesicht. Mault auf eine sachlich gestellte Frage einfach zurück. 😡😡😡"

Graph 5: Average likelihood that posts contain extremist content
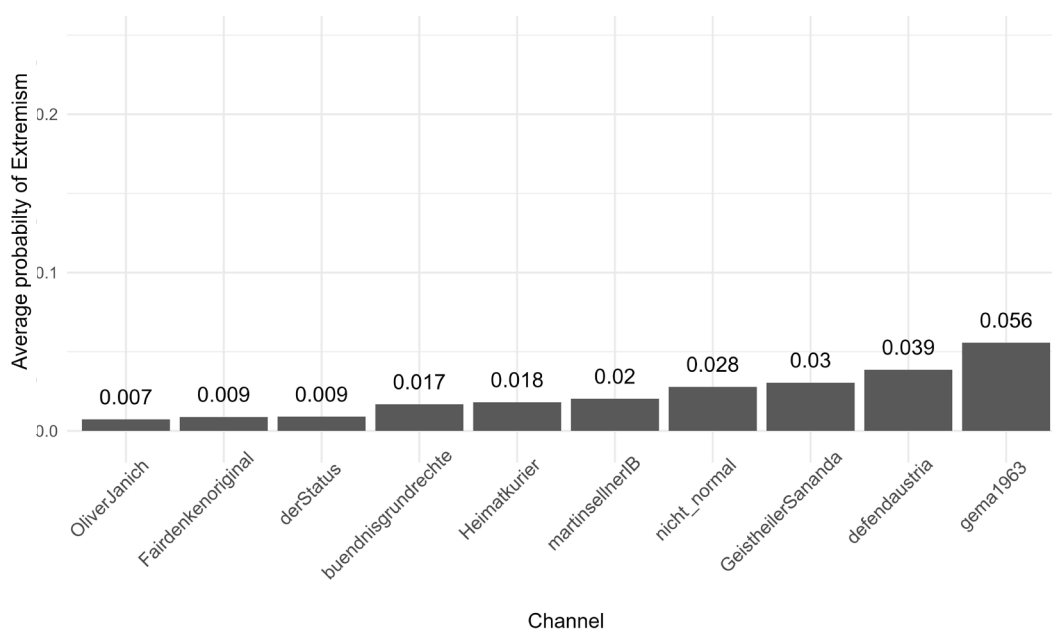
Graph 5 shows the average likelihood that posts contain extremist content across the ten Telegram channels with the highest probabilities. Interestingly, while these channels show a relatively higher average probability of extremism compared to the others, their extremism probabilities were generally lower than the probabilities of hate speech or toxicity for the same channels. This pattern suggests that although extremist content is present among the channels examined, it is not as prevalent as toxic or hate speech content.

The example below, from **martinsellnerIB**, is highlighted as particularly high in extremism probability, which thematically aligns with this channel's content style or the themes it promotes. This example helps to underscore the types of content that the model identifies as extremist, suggesting that **martinsellnerIB** and similar channels might frequently offer extreme ideologies or narratives.

Example 2: Extremist post from channel *martinsellnerIB*, English translation

martinsellnerIB

'🇨🇭 Islammigrantin fires 20 times at the icon of the Virgin Mary 🟥 This disgusting person 'fled' from Bosnia to Switzerland. Born a Muslim, she gets stuck, obtains a passport, doesn't return home, but sticks to her foreign, aggressive ideology. One day, she fires a gun at a central icon of the religion that has characterised her host country for 1,600 years. 👉 I've never heard of this Islamic migrant before, but in view of the constant Islamic terror attacks in Europe, it seems extremely worrying and suspicious to me. 👜 I think: 29 years of #sanjaameti in Switzerland is enough. If this subject has a shred of decency, she will leave the country that has offered her so much and whose identity she has desecrated to such an extent. ❗ Der State still has no means of dealing with this. Her more than 20 shots at the icon are wake-up calls: Switzerland urgently needs to revise its citizenship law and make it possible to revoke citizenship in blatant cases. In any case, I never want to see 'Sanja' again, and I don't want to hear anything more about her in the German-speaking world. How are you? 🎮 This channel will only grow if you share this link: https://t.me/martinsellnerIB'

```
[8] "CHIslammigrantin feuert 20mal auf Marienikone\n\n🟥 Diese widerliche Person \"flüchtet\" aus
Bosnien in die Schweiz. Der geborene Moslem setzt sich fest, erschleicht sich einen Pass, kehrt ni
cht heim, aber hält an ihrer fremden, aggressiven Ideologie fest. Eines Tages feuert sie mit einer
Waffe auf eine zentrale Ikone der Religion, die in ihrem Gastland seit 1600 Jahren prägend ist. \n
\n👉 Ich habe vorher noch nie von dieser Islammigrantin gehört, aber auf mich wirkt das angesichts
ständiger Islamterroranschläge in Europa extrem bedenklich und verdächtig.\n\n🔒 Ich finde: 29 Jah
re #sanjaameti in der Schweiz sind genug. Wenn dieses Subjekt einen Funken Anstand hat, verlässt si
e das Land, das ihr so viel geboten und dessen Identität sie derart geschändet hat.\n\n❗Der Staat
hat hier noch keine Handhabe. Ihre über 20 Schüsse auf die Ikone sind Weckrufe: Die Schweiz muss d
as Staatsbürgerrecht dringend überarbeiten und in krassen Fällen Aberkennungen ermöglichen. Ich jed
en Fall will \"Sanja\" nie wieder sehen, und im deutschen Sprachraum nichts mehr von ihr hören. Wie
geht es euch?\n\n🎮 Dieser Kanal wächst nur, wenn du diesen Link verbreitest: https://t.me/martin
sellnerIB"
```

The overlap of high probabilities across hate speech, toxicity, and extremism on many of the channels implies that certain channels are hotspots for multiple types of problematic content. The lower extremism scores indicate, however, that while inflammatory or hostile language (toxicity) and content targeting specific groups (hate speech) were relatively common, outright extremist discourse — characterised by content that promotes extreme ideologies, radicalism, or violence — is less frequent. The channels **gema 1963, OliverJanich, nicht_normal, GeistheilerSandra,** and **defendaustria** stand out with comparatively high scores across all the three categories of problematic content. Oliver Janich is a known German conspiracy theorist. The originators of the other channels remain unknown.

Recommendations

Based on the above monitoring findings related to selected Telegram channels ahead of the Austrian 2024 general elections, several key themes emerge for further investigation, including:

/ A long-term analysis, to investigate how the trends detected during the election period change over time.

/ A deeper investigation into the three most prolific channels – Eva Herrman

Offiziel, Oliver Janich, and Uncut_News – the topics addressed, and the target audience. Oliver Janich and Uncut_News would be of particular interest, as they show up among the top 10 for hate speech and for toxicity (and also extremism for Oliver Janich). It would also be good, however, to contrast this with the Eva_Herrman_Offiziel channel, to understand what the drivers were for the large volume of content in each case.

/ Contrast the above channels with those, such as auf1tv, that peak for one type of content, but not for others.

/ Further investigation into channels with consistently high likelihoods of hate, toxic, and extremist content.

/ For further analysis, the model results could also be examined to determine who is being targeted if hate speech is detected (that is, whether it is directed against an individual, a group, or the general public). This could similarly be done with the Expression Detector, to determine whether hate speech is explicit or implicit (e.g., implicit hate speech might suggest irony, metaphors, etc., and is generally more difficult to recognise).

/ Finally, from a technical perspective, model effects could also be investigated, to determine whether the models themselves are more likely than others to detect certain kinds of hate or toxic speech.

Date: November 2024

Acknowledgments

ELECTION-WATCH.EU

**DEMOCRACY
REPORTING
INTERNATIONAL**

Annex

The below table shows the performance of different models on tasks such as identifying extremism, toxicity, and hate speech. The metrics include accuracy, precision, recall, and F1-score, reported separately for the validation and test datasets.

Model	Test Accuracy	Test Precision	Test Recall	Test F1-Score
Extremism	0.89	0.71	0.62	0.65
Toxicity	0.78	0.75	0.73	0.74
Hate speech	0.89	0.71	0.65	0.67

Accuracy: Measures the overall correctness of the model, calculated as the number of correct predictions out of all predictions.

Precision: Indicates how many of the items predicted as positive (e.g., engaging comments or hate speech) are actually positive. This reflects the model's ability to avoid false positives.

Recall: Shows how many of the actual positives were correctly identified by the model, reflecting its ability to avoid false negatives.

F1-Score: The harmonic mean of precision and recall, providing a balanced measure of the model's accuracy; especially useful when classes (categories of data) are imbalanced.

Each row represents a specific model task, and the values in each row provide insight into how well the model performs in identifying that particular aspect, both during validation and in a real-world test setting. Higher scores indicate better performance.

Extremism and hate speech show high accuracy, but lower precision and recall on the test set, indicating potential overfitting or difficulty in accurately identifying these categories in varied data.

This analysis helps determine which tasks the model performs reliably and where improvements might be needed.