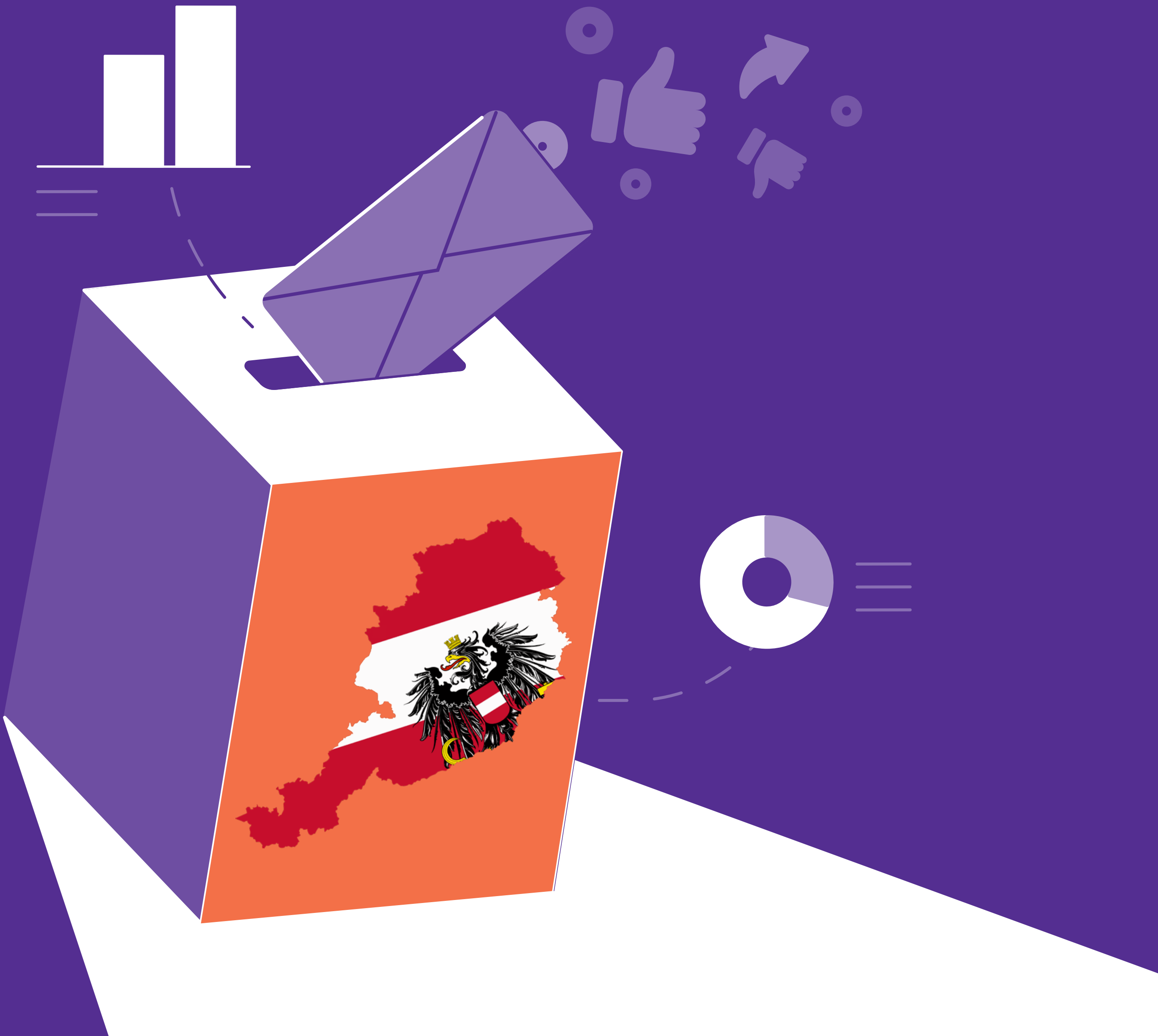
From Hashtags to Votes:

Social Media Patterns in Austria's 2024 National Elections



From Hashtags to Votes:

Social Media Patterns in Austria’s 2024 National Elections

About Wahlbeobachtung.org

Wahlbeobachtung.org is an independent, non-partisan Civil Society Organisation (CSO, association (Verein) under Austrian law) of election experts with international observation and technical assistance experience. Its objective is to contributing to the improvement and strengthening of the Austrian electoral system and electoral processes in particular and the political system in general.

About Democracy Reporting International

DRI is an independent organisation dedicated to promoting democracy worldwide. We believe that people are active participants in public life, not subjects of their governments. Our work centres on analysis, reporting, and capacity-building. For this, we are guided by the democratic and human rights obligations enshrined in international law. Headquartered in Berlin, DRI has offices in Lebanon, Libya, Myanmar, Pakistan, Sri Lanka, Tunisia, and Ukraine.

Acknowledgments

The analysis for this report was conducted by Klara Pernsteiner, from the University of Vienna, and Armin Rabitsch, from wahlbeobachtung.org, with contributions from a senior data scientist and two researchers from an Austrian research institute, whose names remain confidential to ensure their safety.

Forset designed the layout of this publication.

Date: January 2025

This brief is part of the access://democracy project, implemented by Democracy Reporting International and funded by the Mercator Foundation. Its contents do not necessarily represent the position of the Mercator Foundation.

DEMOCRACY REPORTING INTERNATIONAL

access democracy

ELECTION-WATCH.EU

Contents

1. Background and Key Findings

a. Background

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 far right populist Freedom Party (FPÖ) became the biggest party in the lower house of Parliament. The 28.8 per cent of valid votes cast the party received resulted in 57 (31.1 per cent) of the seats in the lower house, the 183-member National Council.[1]

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[2] 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 other 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 election integrity.

Media and state services did not report any detected foreign or party-orchestrated disinformation campaigns, and the 2024 electoral campaign was generally assessed as fair.[3] However, AUF1, labelled by Austrian domestic intelligence as an "alternative right-wing extremist" media channel, alleged election fraud risks, claiming postal voting could be used to prevent an FPÖ victory. It further suggested the existence of 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.

In this study, we examine toxicity, hate speech, and extremism on Austrian online and social media platforms, as well as on Telegram channels, ahead of the elections. This study displays findings in graphs to provide a better understanding and, in addition, includes a disinformation case study.

[1] 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. Three-party coalition talks between the ÖVP, SPÖ, and the liberal NEOS party commenced after the elections.

[2] Major X influencers with hundreds of thousands of followers, such as public broadcaster (ORF) anchor Armin Wolf, or the head of the political weekly news magazine Falter, Florian Klenk, left the Austrian X platform, in an orchestrated move to Blue Sky ,following the involvement of X owner Musk in the US presidential elections.

[3] Austrian and German media did not report 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/

b. Key Findings

- **Thematic nature of the offensive discourse:** Discussions about immigration or ethnicity serve as a common vector for toxic, hateful, or extremist rhetoric. Controversial or emotionally charged topics such as migration often provoke polarised reactions and provide fertile ground for hate speech, and align with broader patterns observed in social media studies. Such findings underscore the importance of addressing racist and xenophobic themes on online platforms to foster a safer and more inclusive digital environment.

- **Domains hosting offensive content:** A small number of highly active online domains are central to discussions on platforms like www.derstandard.com, www.youtube.com, and www.facebook.com, including both offensive and non-offensive content. Of the 408 domains analysed, 84 had at least ten posts or comments, and a smaller subset hosted more than 50 entries. Offensive content, and particularly content labelled as hate speech, toxic, or extremist, receives significantly higher views and reactions compared to non-offensive material. A word cloud analysis highlights how terms like "foreigners" and "migrants" appear disproportionately more often in offensive content, indicating that much of the problematic material focuses on xenophobic or racist narratives.

- **Disinformation narratives of election fraud:** A case study on disinformation surrounding alleged election fraud via postal voting highlights the role of alternative media platforms, particularly AUF1, in spreading unverified claims. AUF1 is one of the largest Austrian extreme-right news channels, which also disseminates its content via a Telegram channel. Despite an increase in absentee ballots over the previous election, no evidence of fraud was found. Analysis of social media trends showed a significant surge in discussions about election fraud following AUF1's coverage, establishing the platform as a key amplifier of disinformation. Notably, this disinformation was not disseminated through paid social media ads, indicating an organic spread through networked discussions.

- **Role of alternative media and conspiracy theorists:** Based on our analysis, we found that the narrative of election fraud was first introduced by conspiracy theorist Martin Rutter,[4] who mobilised protests and disseminated false claims through Telegram. These

[4] See: Blaise Gauquelin & Katharina Zwins, "Austria's far right woos Anti-vaxxers with fund for vaccine 'victims'", Barron's, 23 September 2024.

activities, coupled with AUF1's online media coverage, magnified the disinformation's reach. The association of Austrian discussions with the German Brandenburg elections highlights cross-border thematic resonances in such disinformation.

- **Key Telegram channels:** The channels EvaHermanOffiziell, OliverJanich, 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 the circulation of a substantial portion of this content.

- **Prevalence patterns:** Toxic speech was the most common type of problematic content on the Telegram channels, followed by hate speech, and then extremism, which appeared less frequently. This suggests that the use of inflammatory and offensive language was widespread, while extreme ideological content was comparatively rare.

- **Influence of prominent extremists on Telegram:** Posts with high hate speech, toxicity, and extremism probabilities (e.g., those from Martin Sellner, an Austrian far right extremist, on his Telegram channel martinsellnerIB)[5] 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.

[5] Martin Sellner is the former leader of the Identitarian Movement Austria, which received donations from the gunman in the shootings at two mosques that killed 51 people in Christchurch, New Zealand in 2019. Since 2021, the display of symbols and gestures of the Identitarian Movement has been prohibited in Austria.

2. Methodology

The data for this project was collected and analysed by Austrian security researchers and wahlbeobachtung.org. To ensure broad coverage of the online and social media activities surrounding the Austrian National elections, we combined data from two different sources. First, we relied on online and social media from a wide range of different domains, collected by the third-party platform SentiOne. This data includes content from platforms such as Facebook, X, and YouTube posted between 1 September and 14 October 2024, encompassing the most intense phase of the campaign period leading up to and the two weeks following the 29 September elections. The second part of the data is from 27 Telegram channels, selected based on previous research.

The online and social media content acquired via SentiOne was selected using over 150 keywords related to the elections. These keywords include the names of competing political parties and their top candidates, as well as general election-related terms. Posts from 408 Austrian domains were included, regardless of whether the content explicitly referenced the Austrian context. For content from foreign domains (e.g., ".com" or ".de"), inclusion was limited to posts that clearly related to the Austrian elections. This was achieved by applying domain-specific keyword filters. The keyword and domain selection process ensured that only posts and comments specifically relevant to the Austrian elections were included in the analysis. SentiOne does not have access to all domains and social media platforms – private Instagram accounts are not accessible, for example. We had to accept the legal and practical limitations of the service SentiOne offers, as it was the best available option for gathering such a wide range of online data. This led us to a total of 17,011 observations, including both posts and comments, on various platforms and in various domains.

The 27 Telegram channels were selected based on existing research and information 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.[6] Furthermore, the channels were selected because the above-mentioned reports concluded that these were the most influential in Austrian political public discourse and, therefore, were 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. 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). In total, combining data from SentiOne and Telegram left us with 26,324 observations to analyse.

Both data sources were analysed with a focus on detecting extremism, hate speech, and toxicity, using machine learning classifiers developed by the researchers.

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

While “toxicity”, “hate speech”, and “extremism” are closely related, they are not interchangeable, and can occur independently of each other. To distinguish between the three categories, the following definitions were used for the automated detection models:

- **Toxicity** indicates a comment’s potential 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 based on characteristics attributed to the groups.
- **Extremism** is any form of extreme or radical (political or religious) statement that is at odds with the democratic order.

Toxicity, hate speech, and extremism detection tools were used to analyse the social media content. The basis of the detection tools is XLM-RoBERTa,[7] a high-performing language model that is trained on multilingual data. The researchers 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.

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

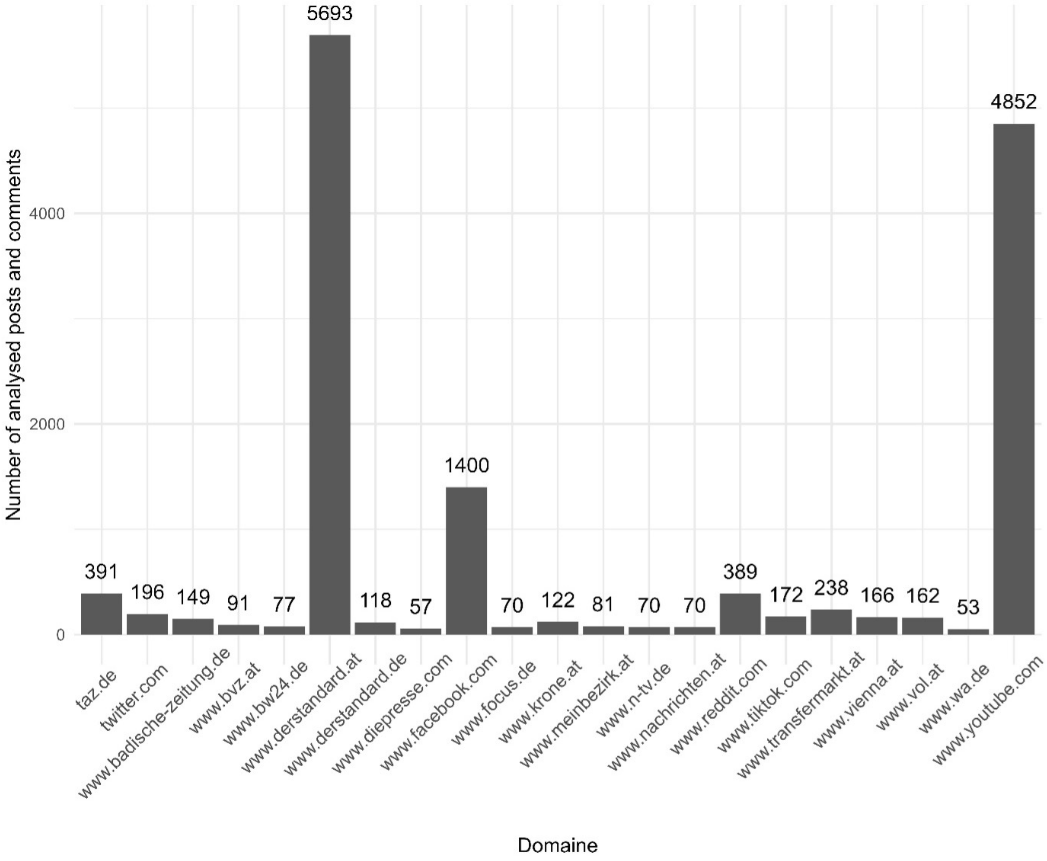
3. Findings

a. Findings on SentiOne

The following analysis provides answers to the overarching question of to what extent different forms of offensive content were present in the online political discourse surrounding the Austrian national elections. To answer this question, we combined univariate and bivariate descriptive statistics and machine learning tools with the above described data sources, along with conducting a case study.

The three graphs below offer a detailed picture of how offensive content spread and engaged audiences, emphasising the need for nuanced strategies to curb its influence. They highlight the importance of understanding not just where such content originates, but also how it interacts with platform dynamics and user behaviour.

Graph 1: Domains with More than 50 Posts or Comments

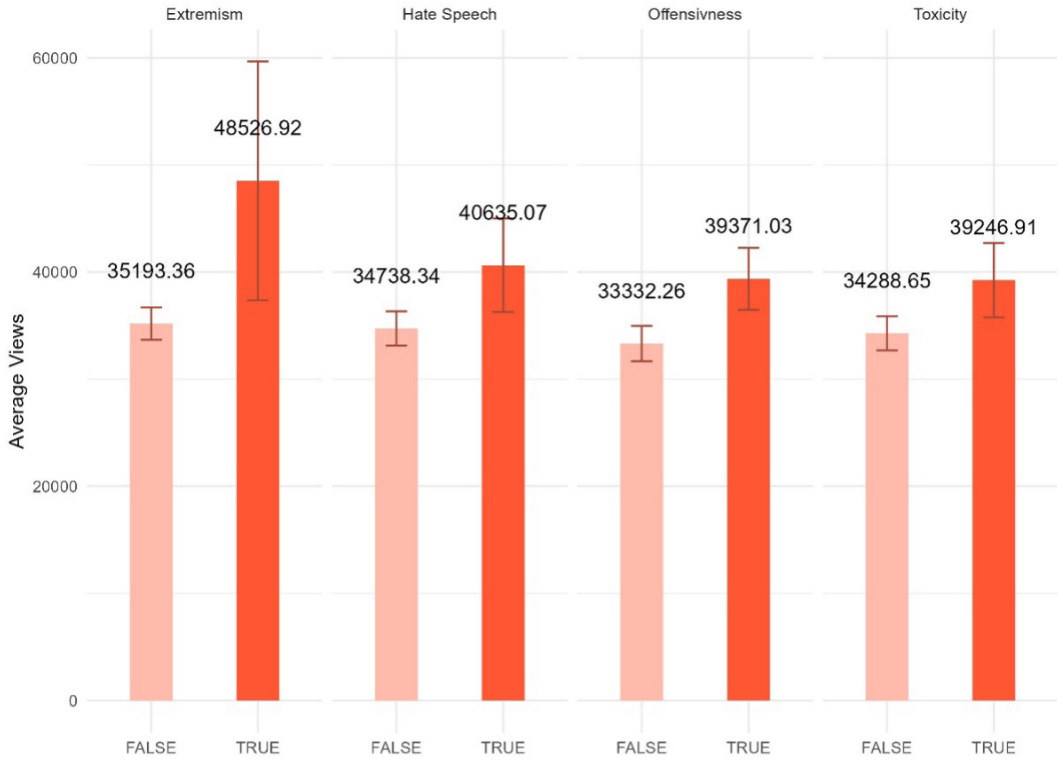


Graph 1, above, highlights the domains contributing the largest number of posts and comments in the dataset, focusing on those with over 50 entries. These domains represent the most active platforms in terms of discussions, which include both offensive and non-offensive content. The significance of these domains lies in their role as hubs for online engagement, making them critical spaces for monitoring and addressing the spread of offensive material.

In total, the data set includes 17,011 posts and comments from 408 different domains. For the analyses, however, only those platforms with at least ten comments detected were included. Thus, the SentiOne data analysis covers 84 platforms. The three domains with the most content included are www.derstandard.com, www.youtube.com, and www.facebook.com. The first is the website of one of the largest Austrian daily newspapers and has a very active community forum, which often hosts political discussions. It is not surprising, therefore, that together with the two popular social media platforms Facebook and YouTube, it is the largest contributor of content.

Content was labeled as hate speech, toxic, or extremist if it had a respective probability higher than 50 per cent. Furthermore, content was labeled as offensive if it contained at least one of the three types of problematic content. We compared non-offensive vs offensive content. This can be seen in Graph 2 for views, and in Graph 3 for reactions (likes and comments). Reactions combine the different types of possible engagement and, therefore, offer a more nuanced insight into how much attention the different types of content attracted. The analysis showed that each kind of offensive content had a significantly higher number of average views than non-offensive content. This confirmed the assumption that problematic content spreads faster than non-problematic content.

Graph 2: Comparison of Views Between Non-Offensive and Offensive Content

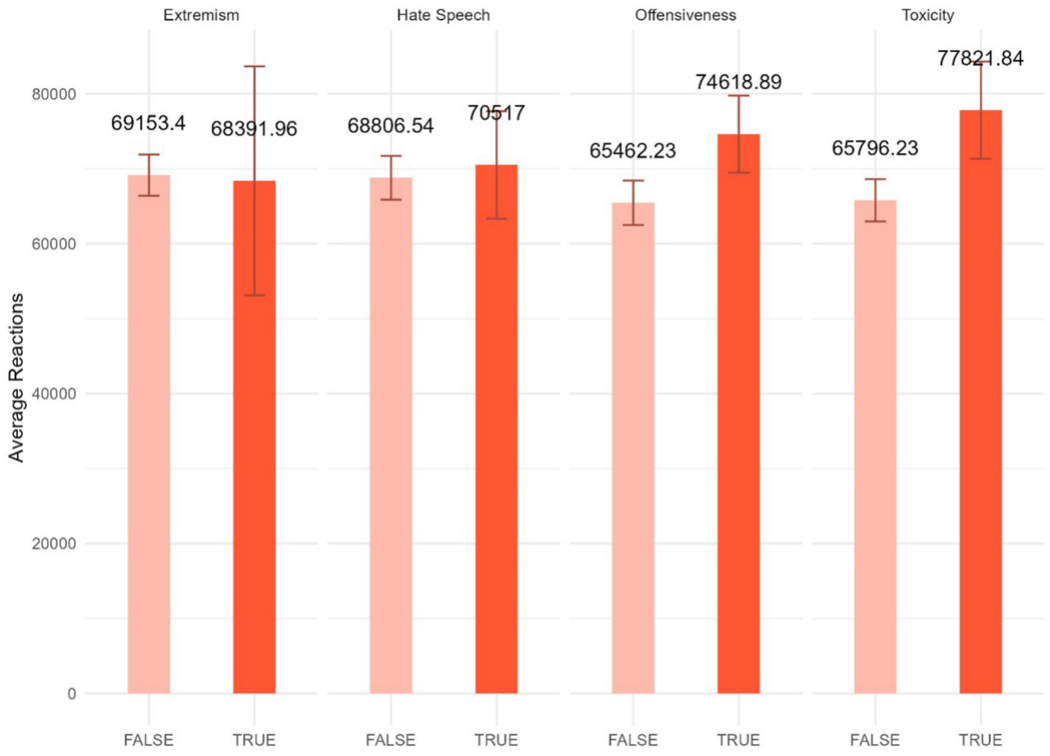


Graph 2 demonstrates that offensive content consistently attracted more views than non-offensive content. This difference suggests that offensive material is inherently more engaging or visible to audiences. Several factors could contribute to this phenomenon:

1. **Algorithmic amplification:** Many platforms use algorithms that prioritise engagement, potentially pushing controversial or provocative content higher in users' feeds.
2. **Human curiosity:** Offensive content, particularly material classified as hate speech or toxic, often sparks curiosity, outrage, or shock, which might drive higher click-through rates.
3. **Viral potential:** The divisive nature of offensive content can make it more likely to be shared, discussed, and debated, further increasing its visibility.

This finding underscores a critical challenge for platforms – while engagement metrics drive content visibility, they can also inadvertently promote harmful material. Solutions to this problem might include revising algorithms to down-rank offensive content, or improving detection systems to flag and remove such material more quickly. Article 34 of the EU's Digital Services Act (DSA) provides for very large online platforms (VLOPs) and very large online search engines (VLOSEs) to diligently identify, analyse, and assess any systemic risks from the design and functioning of their services, including algorithmic systems, or from the use of their services. This could mean that VLOPs and VLOSEs would need to adjust the algorithms to avoid offensive material resulting in more engagement, and that the European Commission will request further transparency measures regarding content moderation, or even issue fines if VLOPs and VLOSEs do not comply. Election-Watch.EU continues to recommend the full transparency of VLOPs and VLOSEs algorithms for third party scrutiny, to avoid bias and the amplification of hateful and extremist content, and to bring public spaces increasingly back under democratic oversight.

Graph 3: Reactions to Non-Offensive vs Offensive Content

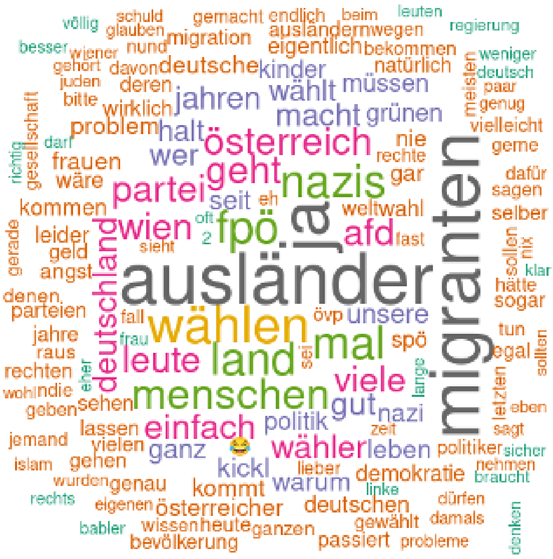


The third graph, above, examines reactions, which include likes, comments, and views. Comments are counted twice, as they constitute a more intensive form of interaction than views or likes. Graph 3 builds on the findings of Graph 2 by exploring how users actively engage with content. While offensive content statistically generated significantly more reactions overall, some nuances emerged:

- **Toxic content:** This category drove the most significant increase in reactions, suggesting it was particularly effective at eliciting user interaction. Toxic content often thrives on provocation, leading to heated discussions or debates.
- **Hate speech and extremist content:** Unlike toxic content, these categories did not show a significant difference in reactions compared to non-hate speech and non-extremist content. This might indicate that, while hate speech and extremism can garner views, they are less likely to provoke interactive engagement, such as likes or comments.

This difference in engagement patterns is important because it highlights the varied dynamics of different types of offensive content. Toxic content may be more emotionally charged, prompting users to react immediately, whereas hate speech or extremist content might provoke passive consumption, rather than active interaction.

Graph 4: Word cloud



Graph 4 presents a word cloud that visualises terms that appear as among the 100 most frequent words used in offensive content. The size of each word in the cloud corresponds to how frequently it is used in offensive posts, with larger words indicating a higher prevalence. The two most prominent terms – "foreigners" (“Ausländer”) and "migrants" (“Migranten”) – stand out, suggesting a substantial focus on topics related to immigration and ethnicity in the offensive material. This strongly implies that a significant portion of the offensive content was driven by racist or xenophobic narratives.

Such content likely targeted specific groups or communities based on their ethnic or national origin, perpetuating stereotypes, hate speech, and/or discriminatory attitudes.

To further investigate the meaning of the content, we use a model with k=10 topics. Table 1 in Appendix 1 shows the ten most important words for each topic, as well as the topic's frequency and the average offensiveness. The model shows that topic 9, labelled as right-wing populism, is the topic with the highest average probability of offensiveness, and also the most common topic. For the most offensive and most common topic, words like "simple" ("einfach"), "people" ("Menschen"), and "foreigners" ("Ausländer") point to the populist and anti-immigrant nature of the offensive content. The topic with the second highest probability of offensive posts, labelled election results and the FPÖ top candidate, includes mostly keywords related to parties and the top-candidate of the far right populist FPÖ. This indicates that content that included terms that are very generally related to the election are also used in content that is offensive in nature. Other topics include German politics, US elections, the Russian war in Ukraine, election results, and coalition options. Additionally, the results of this topic model confirmed previous findings that offensive content attracts more views than non-offensive content.

Disinformation case study: Alleged election fraud via postal voting

Leading up to the election, so-called "alternative media" reported on potential election fraud regarding postal voting. The media channel AUF1 stated that the record high number of postal ballots could be used "to steal the election" from the far right FPÖ and its leader, Herbert Kickl. While it is true that the number of postal ballots issued for this election was significantly higher than in the last national election, in 2019 (1,436,240 such ballots in 2024, vs. 1,070,933 in 2019), there was no indication of election fraud and no party lodged any official complaints. In both the original AUF1 article and in subsequent discussions, this accusation of election fraud was mentioned in association with the German regional elections in Brandenburg, where the Alternative für Deutschland (AfD) party ultimately lost its lead after the count of postal ballots.[8] Accusations related to postal ballots are a standard part of a right-wing populist playbook, copied from campaigns such as those of Donald Trump in the United States.[9]

The first known mention of the potential election fraud was by Rutter, the conspiracy theorist and former Austrian politician. He registered a demonstration "against election fraud" on 2 September, four weeks before the election. Subsequently, he regularly posted on his telegram channel that "only

[8] See: Correctiv; Faktencheck; Landtagswahl Brandenburg: Unterschiede zwischen Brief- und Urnenwahlergebnis belegen keinen Betrug; 18 October 2024.

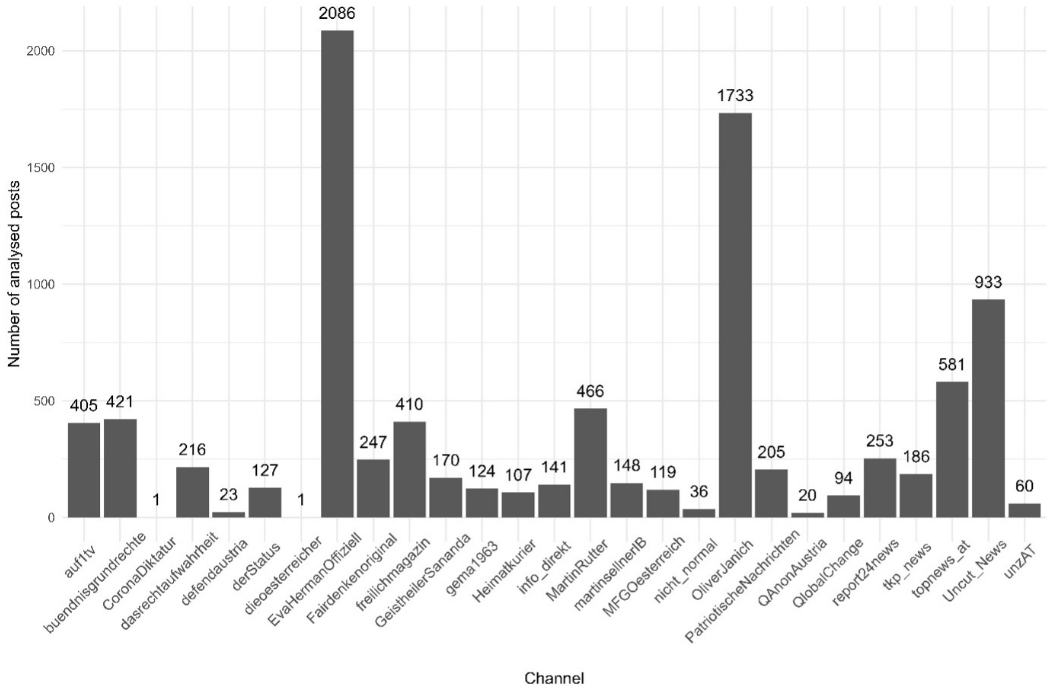
[9] Such as Mike Wendling, "Whirlwind of misinformation sows distrust ahead of US election day", BBC, 3 November 2024.

large scale election fraud could stop the victory of the FPÖ”. The first AUF1 article claiming that there was a danger of election fraud was published on 24 September. Wahlbeobachtung.org social media analysis in the period from 2 September to 6 October showed that the number of people reached by posts mentioning election fraud in association with the national elections surged significantly after 24 September. This is a strong indicator that the alternative media platform AUF1 is, at least to some extent, the relevant multiplier of this disinformation, which is then spread further via social media. Based on findings from the Meta Ad library, there was no indication that this disinformation was spread via paid ads on social media.

b. Findings on Telegram

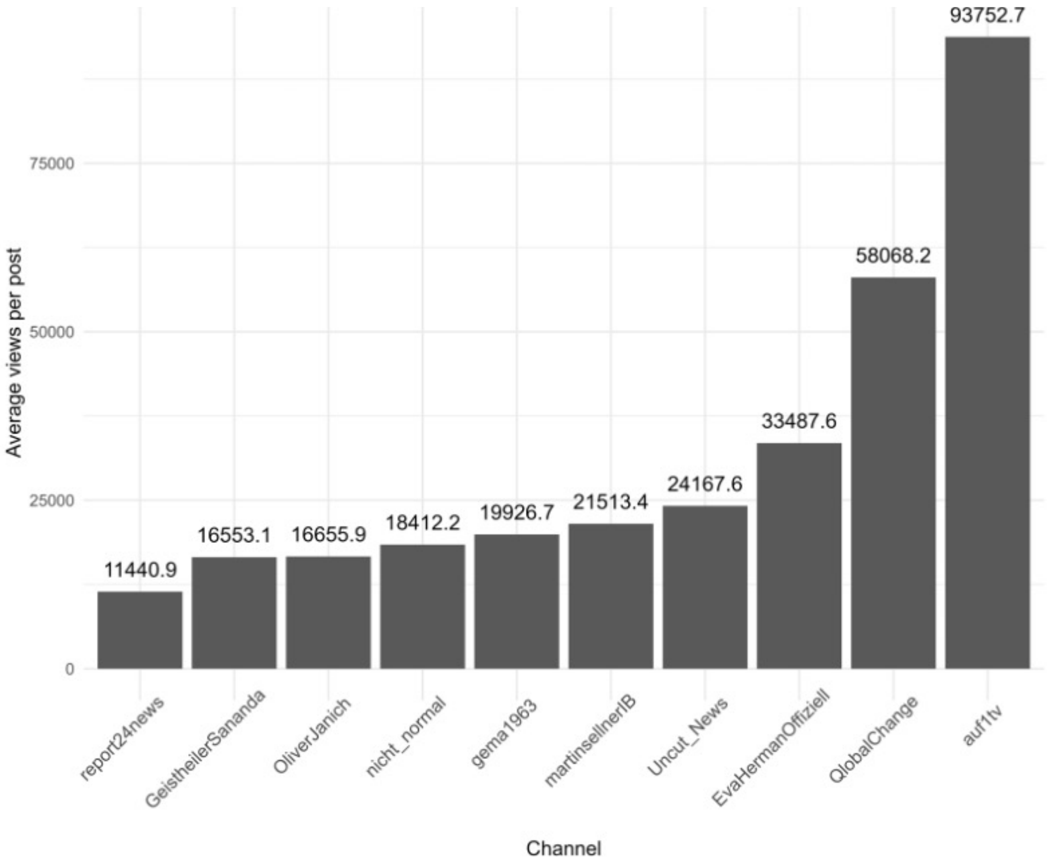
Below, we show the prevalence of hate speech, toxicity, and extremism on the different Telegram channels. To illustrate this further, we give examples of posts (translated) that had particularly high scores on the respective scales. The analysis shows that there were significant differences both across 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, nor that we should be unconcerned about the distribution of extremist content via Telegram.

Graph 5: Overview of All Analysed Telegram Channels, and the Volume of Posts



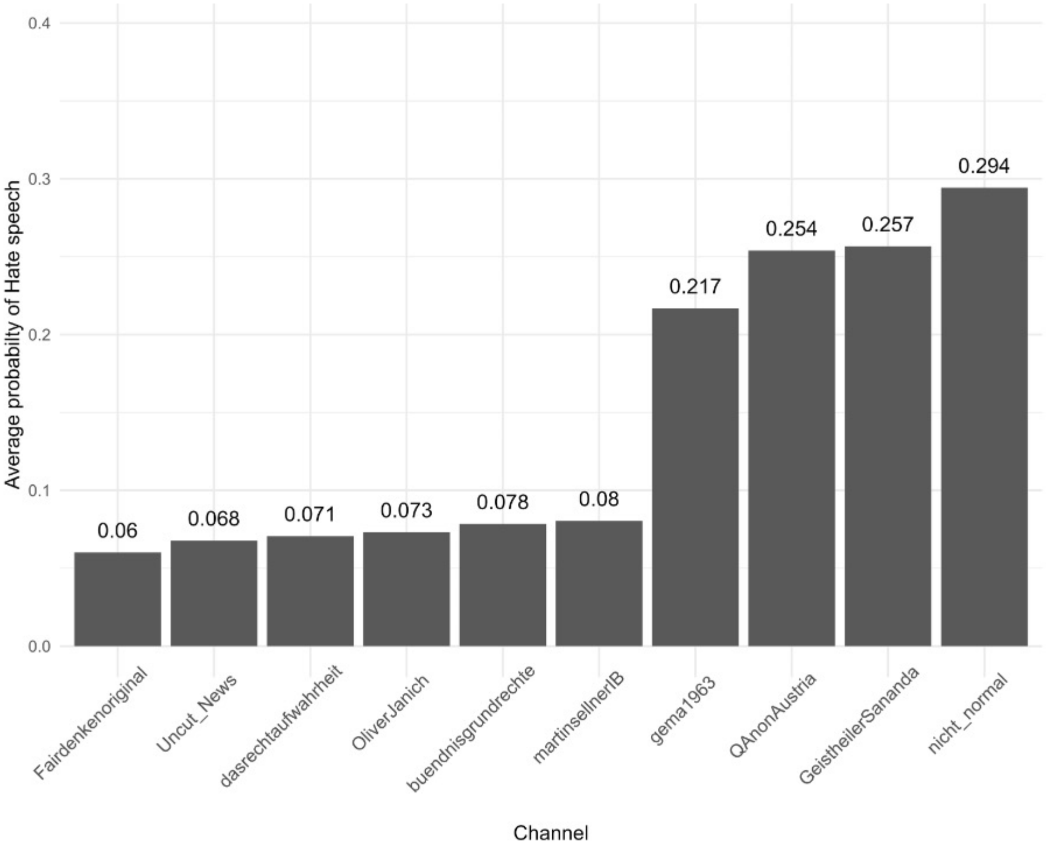
Graph 5 provides an overview of all analysed Telegram channels and the volume of posts included from each channel. The three channels with the highest number of posts — EvaHermanOffiziell, OliverJanich, and Uncut_News — stand out as particularly active. Their elevated post counts, together with the high rate of views per post, suggest they were focal points for communication, and likely influential in the discourse being studied here. 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 have been excluded from the subsequent analyses.

Graph 6: Average Views per Post



Graph 6 shows the average views per post for the ten channels with the most views. The views per post provide an 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 impact of this content within and beyond Telegram.[10] The channels auf1tv, QlobalChange, and EvaHermanOffiziell had the highest number of average views in our sample. Neither QlobalChange nor EvaHermanOffiziell 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 7: Illustration of the Distribution and Likelihood of Hate Speech Content

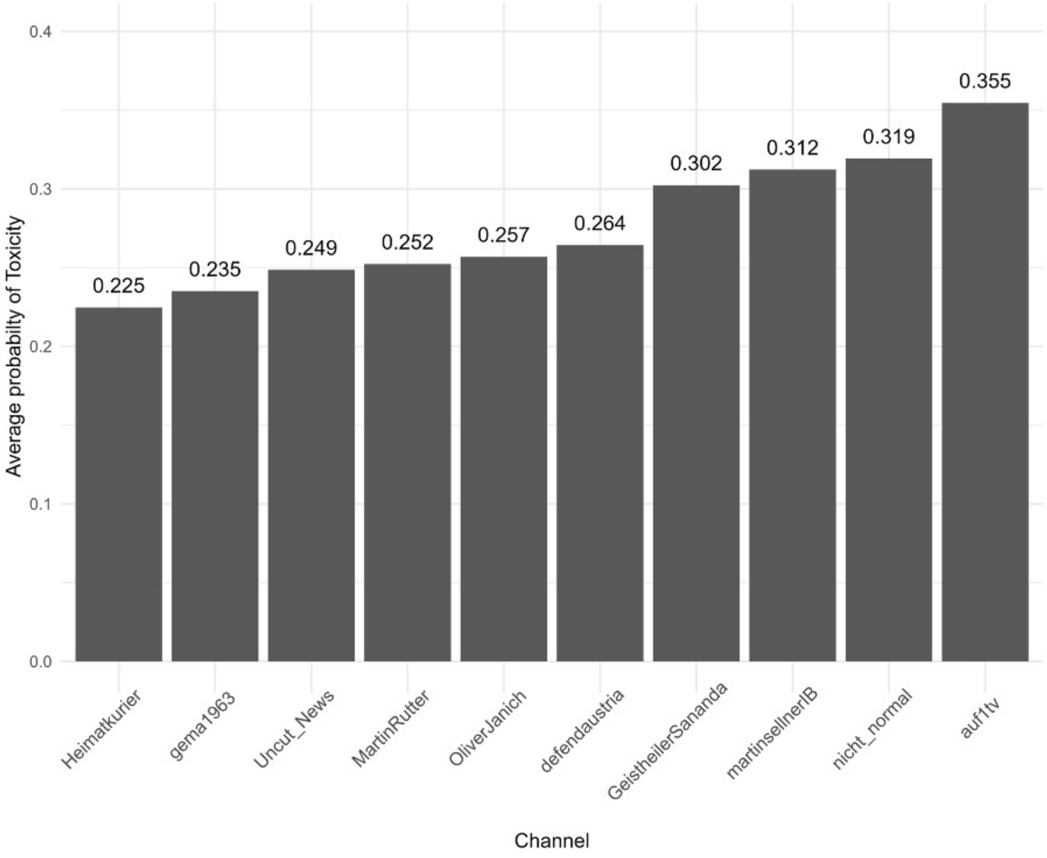


Graph 7 provides an illustration of the distribution and likelihood of hate speech content on some of the observed Telegram channels, focussing on the top ten. Channels with consistently higher probabilities suggest a sustained tendency towards the use of hate speech, making them critical

[10] 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.

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), OliverJanich and Uncut_News, were also among the top ten for the likelihood of hateful content. This finding suggests a need for further investigation into these channels, and specifically into the types of hateful content they disseminate, as well as the targets thereof.

Graph 8: Average Probability that a Post Includes Toxicity for the Ten Channels with the Highest Average Probability

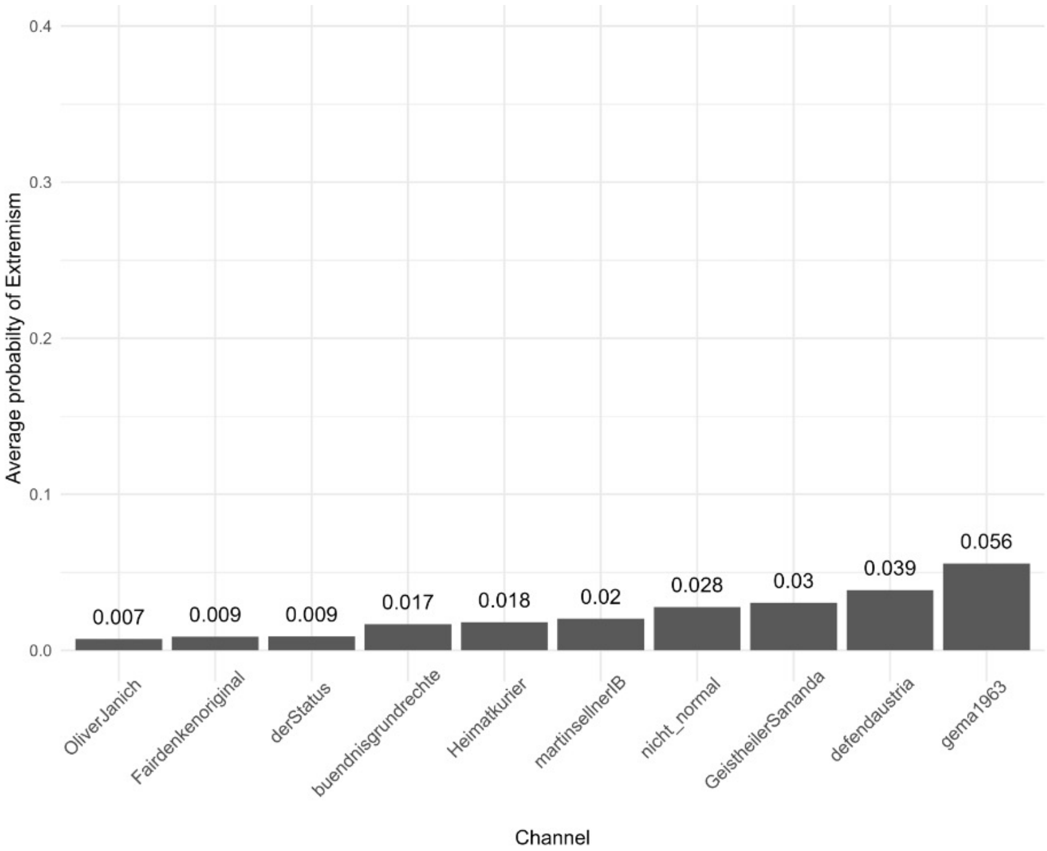


Graph 8 shows the top ten channels in terms of average probability that a post included toxicity. Of particular note here is the presence of auf1TV, with the highest average likelihood of toxic content in a post, since the channel is not present at all in the top ten for hate speech or extremist content. 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, from the channel buendnisgrundrechte. In this post, a conservative politician is called a slut and accused of lying.

Example 1: Toxic Post from the Channel buendnisgrundrechte

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

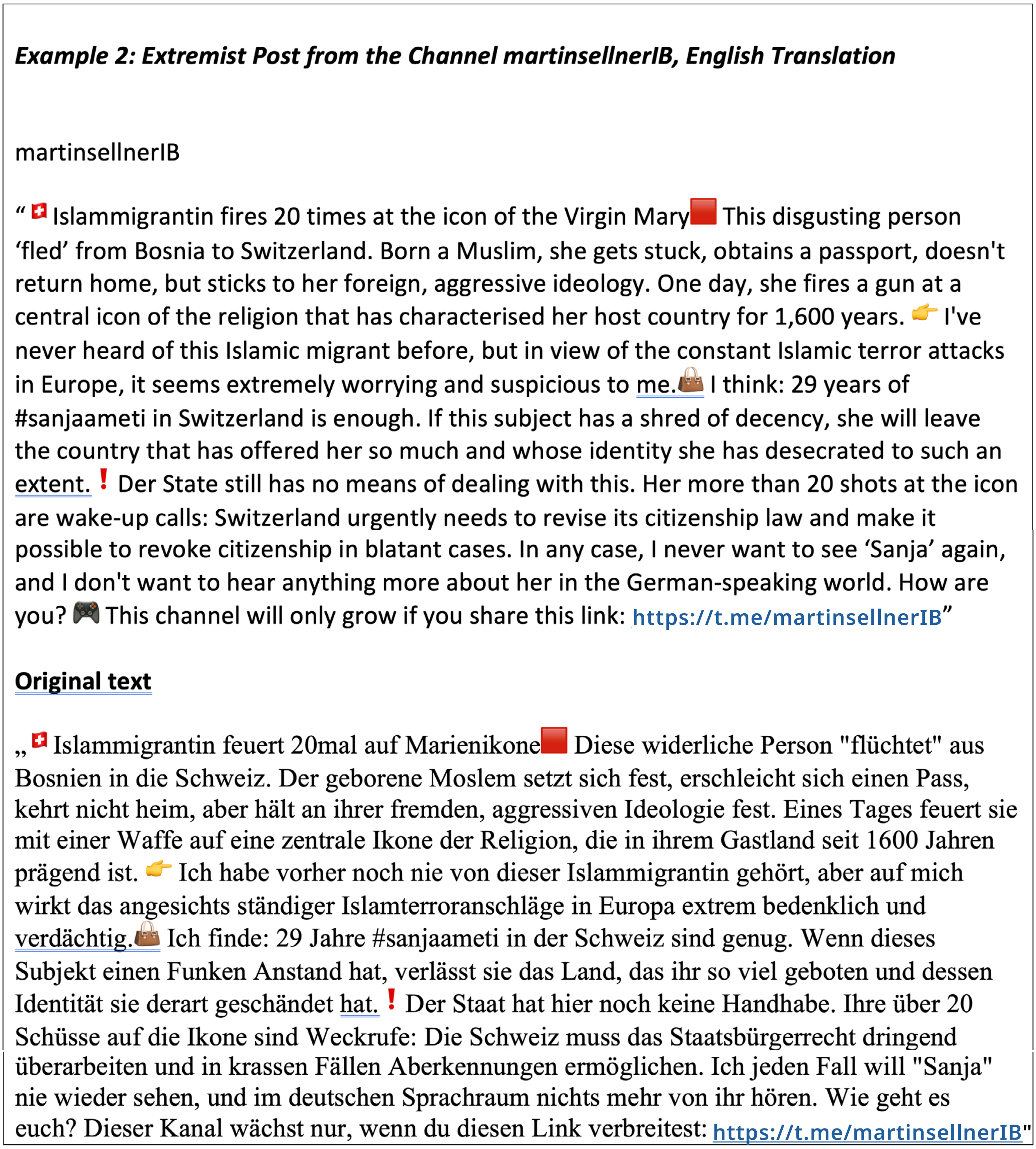
Graph 9: Average Likelihood that Posts Contain Extremist Content



Graph 9 shows the average likelihood that posts contained 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 the probability

> of extremist content, which aligns thematically 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 promote extreme ideologies or narratives.



Example 2: Extremist Post from the 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”

Original text

„ 🇨🇭 Islammigrantin feuert 20mal auf Marienikone🟥 Diese widerliche Person "flüchtet" aus Bosnien in die Schweiz. Der geborene Moslem setzt sich fest, erschleicht sich einen Pass, kehrt nicht 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. 👉 Ich habe vorher noch nie von dieser Islammigrantin gehört, aber auf mich wirkt das angesichts ständiger Islamterroranschläge in Europa extrem bedenklich und verdächtig.👜 Ich finde: 29 Jahre #sanjaameti in der Schweiz sind genug. Wenn dieses Subjekt einen Funken Anstand hat, verlässt sie das Land, das ihr so viel geboten und dessen Identität sie derart geschändet hat. ❗ Der Staat hat hier noch keine Handhabe. Ihre über 20 Schüsse auf die Ikone sind Weckrufe: Die Schweiz muss das Staatsbürgerrecht dringend überarbeiten und in krassen Fällen Aberkennungen ermöglichen. Ich jeden Fall will "Sanja" nie wieder sehen, und im deutschen Sprachraum nichts mehr von ihr hören. Wie geht es euch? Dieser Kanal wächst nur, wenn du diesen Link verbreitest: https://t.me/martinsellnerIB"

The overlap of high probabilities of 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 well-known German conspiracy theorist. The originators of the other channels remain unknown.

Based on the Telegram monitoring findings, several key themes emerge for further investigation, including a long-term analysis of how the trends detected during the election period change over time. Another would be a deeper investigation into the three most prolific channels – EvaHermanOffiziell, OliverJanich, and Uncut_News. OliverJanich and Uncut_News are of particular interest, as they show up among the top ten for hate speech and for toxicity (for Oliver Janich, extremism as well). It would also be good, however, to contrast this with the EvaHermanOffiziell channel, to understand what the drivers were for the large volume of content in each case. Further research could focus on contrasting the above channels with those, such as auf1tv, that peak for one type of content, but not for others. To better understand channels with consistently high likelihoods of hate, toxic, and extremist content, further analysis could aim at examining model results to determine who is being targeted where 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 detection of expressions 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.

4. Recommendations

1. The European Commission (EC) must actively fulfil its DSA oversight responsibilities for VLOPs and VLOSEs, while the Austrian government, Digital Services Coordinator (DSC), and civil society organisations (CSOs) should advocate for platforms to publish algorithmic impact assessments and risk mitigation reports. These should include measures to down-rank offensive content as part of their systemic risk mitigation strategies.

2. The EC and DSC should establish fully independent third-party audits of algorithms to enhance the detection and prevention of bias. As mandated by the DSA, platforms must contract audit firms, and assess and mitigate systemic risks, including those arising from the algorithmic amplification of harmful content, to ensure compliance and transparency.

3. The DSC and CSOs must prioritise strengthening digital literacy initiatives, particularly targeting younger demographics on platforms such as Instagram, Telegram, and TikTok, aligning these efforts with users' digital habits. This is in line with the DSA's emphasis on supporting public awareness campaigns to counter disinformation and harmful content.

4. The DSC should formally notify platforms of the need to enhance their content moderation and monitoring mechanisms, especially for high-risk channels (e.g., AUF1 and OliverJanich) identified as sources of hate speech. The DSC should also require regular reporting on compliance measures, to ensure accountability.

5. The EC and DSC should collaborate with platforms to integrate advanced multilingual AI models, as encouraged by the DSA. Particular attention should be paid to training these models to recognise and address content in different dialects, argots, jargon, and slang, to improve moderation accuracy.

6. The EC must ensure researchers' access to VLOPs' and VLOSEs' data, in compliance with DSA Article 40. This includes introducing standardised guidelines for platforms to facilitate data access for systemic risk research and enforcing penalties for non-compliance with transparency and data-sharing obligations.

7. The EC and DSC should empower CSOs to contribute to systemic risk assessments, by providing them with the necessary tools, training, and resources to monitor offensive content effectively and report their findings to the DSC and platforms.

5. Annex

a. Topic modelling

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Ascribed label	**German politics**	**US elections and Russian war in Ukraine**	**Austrian and German society**	**Election results and FPÖ top candidate**	**Beer party**
	sei	Trump	Österreich	Kickl	Wien
	grünen	Harris	Menschen	Österreich	Österreich
	sagte	Ukraine	Deutschland	Prozent	Bierpartei
	Merz	Russland	sei	ÖVP	Wlazny
	Partei	sei	Kurz	Partei	sowie
	Deutschland	russischen	džihić	SPÖ	Foto
	Union	USA	AfD	Herbert	ORF
	Söder	russische	Gesellschaft	Nationalratswahl	seit
	müssen	sagte	wäre	Wahl	sei
	Regierung	Donald	bereits	Wien	Wiener
Frequency	5%	4%	3%	5%	5%
Average Offensiveness	14%	14%	17%	25%	14%
Average Reactions	31,504	8,367	10,680	19,296	9,353

	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic NA
Ascribed label	Election results and coalition options	No obvious connection to elections	German regional elections	Right-wing populism	Election forecasts and results	
	ÖVP	Bild	Uhr	ja	Uhr	
	Koalition	60	AfD	AfD	Österreich-wahl	
	SPÖ	💙	Thüringen	Mal	ÖVP	
	Wahl	Uhr	Sachsen	wählen	Kickl	
	Prozent	beim	CDU	Ausländer	update	
	Kickl	zwei	BSW	Menschen	Österreich	
	Österreich	September	ganztägig	einfach	Nehammer	
	NEOS	Menschen	Berlin	viele	ersten	
	Parteien	Oktober	2024	Partei	Herbert	
	2024	1	sagt	Leute	Hochrechnungen	
Frequency	2%	5%	2%	67%	2%	1%
Average Offensiveness	8%	19%	14%	46%	14%	25%
Average Reactions	9,689	30,514	11,919	28,396	10,282	28,554

b. Methodological background

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 values 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.

DEMOCRACY
REPORTING
INTERNATIONAL

access
democracy

ELECTION-WATCH.EU