Funded by the European Union

………

Grant Agreement Number: 101094684

Project name: RECOGNITION AND ACKNOWLEDGEMENT OF INJUSTICE TO STRENGTHEN EQUALITY

Project acronym: RAISE

Call: HORIZON-CL2-2022-TRANSFORMATIONS-01

Topic: HORIZON-CL2-2022-TRANSFORMATIONS-01-08

Type of action: HORIZON Research and Innovation Actions Granting authority: European Research Executive Agency

Project starting date: fixed date: 1 May 2023

Project end date: 30 April 2027

Project duration: 48 months

………

Report on Online Boundary Making in Sports

Deliverable 2.1

Cigdem Kentmen-Cin, Senem Kumova Metin, Hande Aka Uymaz

Izmir University of Economics

Disclaimer:

“Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.”

Dissemination Level: Public

Revision History Log

Revision	Date of release	Author	Summary of changes
0.1 - draft	28-04-2025	CKC, SKM, HAU	Initial draft

Table of Contents

Executive Summary

This deliverable (D2.1) reports the findings of Work Package 2 of the RAISE project. The aim of the study was to explore how people relate to group boundaries in their online communication, and how this contributes to the reproduction of structural racism in the contexts of sports. This involved the analysis of over four million user comments on YouTube videos related to four major international sporting events: UEFA Euro 2020, UEFA Euro 2024, UEFA Women's Euro 2022, and the Paris Olympics 2024. Using Natural Language Processing techniques, the study identified hateful comments, analyzed their frequency and content, and identified the most frequently targeted individuals and social identities. It also examined whether, and if so, how, sports organizations responded to these comments on their official YouTube channels. The key findings include:

- Euro 2020 had the highest number of comments and videos. However, among the four events, the 2024 Paris Olympics recorded the highest proportion of hate-labeled comments. This suggests that the multi-sport nature and global scope of the Olympics intensify national rivalries and in-group/out-group tensions.
- Women's Euro 2022 had the fewest comments overall. However, it not only exhibited the second-highest proportion of hate speech but also had the highest percentage of hate comments with likes. Nevertheless, it also had the lowest average number of likes per hate across all events. This pattern suggests that women's football remains a contested space, but also that public endorsement of hateful views is limited.
- Video comments and hateful language typically increased around the time of high-stakes moments, such as knockout phases and final games, but diminished quickly in the following weeks.
- The highest level of personal targeting was observed during Euro 2024, with athletes being the most frequently targeted individuals, with references to nationality, race, and religion.
- While occurring less frequently, hate was also directed at politicians, musicians, and fans, demonstrating that online hostility extends beyond those directly involved in game outcomes to include peripheral individuals within the sporting ecosystem.
- References to nationality or country were the most common in hate comments across all events, with the exception of Women's Euro 2022, where gender references were more frequent.
- References to England and English identity dominate the country/nationality category, followed by other high-profile contestants such as Italy, France, Spain, China, and Germany in international sport events. Other frequent comments attacked non-white or non-Christian players in European national teams during football championships. Contrary to expectations, Russia and Ukraine received comparatively limited attention in the dataset.
- Commentators referred not only to countries but also to symbolic or politically charged places (for example, Wuhan, Gaza, Tiananmen), often in relation to conspiracy theories regarding COVID-19 or historical grievances.
- The presence of hate speech on official institutional channels demonstrates that even professionally managed digital spaces are vulnerable to discriminatory discourse. None of the sports organizations examined in this report directly responded to hate comments on their YouTube channels

1. Introduction

1.1 Aim and Scope

This research report and factsheet, analyzing the expression of racism and xenophobia in sports-related YouTube content, corresponds to Deliverable 2.1 (D2.1) conducted under Work Package 2 (WP2) of the 'RAISE-Recognition and Acknowledgement of Injustice to Strengthen Equality' project. Using big data, WP2 focuses on formulation of group boundaries in online communication and expression of racism and xenophobia in the domain of sports, with three key objectives: (O2.1) retrieving social media data to understand how users communicate group boundaries online and how this reproduces structural racism; (O2.2) systematically coding and classifying social media data; and (O2.3) performing quantitative analyses to show boundary-making practices in the European context. These objectives were achieved via four interrelated tasks. First, search queries were developed to retrieve relevant YouTube data on sports (T2.2). Second, comments posted on YouTube videos were collected during the full duration of each of the following major international sporting events: the UEFA European Football Championships (Euros) 2020 and 2024, the UEFA Women's Euro 2022, and the Paris Olympics 2024, as well as during the month either side of each tournament. Data were also collected from YouTube videos that were popular in the Netherlands, Belgium, Germany, Hungary, Poland, and Türkiye during the same time period (T2.1). Third, these comments were coded to identify boundary-making language, intended targets, and sports organizations' responses to racist or xenophobic remarks (T2.3). The final stage was the analysis of user discourses and boundary-making practices within the collected dataset (T2.4). This deliverable (D2.1) presents the results of these tasks.

1.2 Background and Rationale

Social media sites such as Facebook, Reddit, TikTok, X (formerly Twitter), and YouTube allow users to interact, promote ideas, products and services, entertain themselves, and create and share information (Aichner et al., 2021; Correa and Jeong 2010; Duncombe, 2019; Levinson et al., 2020; Voorveld et al., 2018). The (re)production of social identities is often at the core of these activities; online user interaction contributes to the "process of constituting and re-configuring groups by defining the boundaries between them" (Wimmer, 2008: 1027). This is especially evident in the context of international sports events (Glynn and Brown, 2022). As Kearns et al. (2023: 420) put it "one of the key characteristics of sport is that by the nature of its competitive scope, different countries and cultures frequently compete as peers and opponents on the world stage, and as a result, fans of sport learn about 'the other' through the experience of following their own team/country". Social media further exacerbates boundary-making practices in sports, frequently manifesting as "significant negative online interaction and in many cases, such abusive and or/threatening discourse" (Kavanagh et al., 2016: 786).

This process occurs through a combination of intentional or unintentional algorithms, moderation policies, and platform governance mechanisms (Collier, 2021; Kaplan et al., 2022; Matamoros-Fernández, 2017; Siapera and Viejo-Otero, 2021). In this process, anonymous and pseudonymous identities on social media reduce user accountability, facilitating the expression of aggressive, impulsive, and hateful content (Burnap and Williams, 2015; Cleland, 2014; Matamoros-Fernández, 2017). Moreover, while traditional media is subject to editorial oversight, social media lacks any such systematic gatekeeping mechanisms to control content before publication (Brown, 2018). Although platforms implement content moderation, scholars often describe their practices as a 'black box' due to the lack of transparency in defining, identifying and mitigating hate speech and racism (Myers, 2018; Reynolds and Hallinan 2024). Moderation responsibilities are also divided among multiple actors, including platform administrators, human moderators, end-users, and algorithmic tools, resulting in inconsistencies in detecting and removing controversial content (Caldevilla-Dominguez et al. 2023; Dubois and Reepschlager, 2024; Einwiller and Kim, 2020; Konikoff, 2021; Matamoros-Fernández, 2017). Within this environment, the emotional intensity of international sports events and sports rivalry often fuel online negative othering, expressed through threatening or degrading comments aimed at racial, ethnic, religious, and gender minorities.

In this work package, we focus on the frequency and patterns of this hateful and offensive othering in sports within comments on the YouTube platform, shown by recent studies to be the world's leading video-sharing platform, regularly used by 85 percent of adults in the United States and 56 percent across the European Union (Dannenbaum, 2025; European Parliament, 2022). What makes YouTube appealing to its millions of users is not only the videos uploaded to customized channels by content creators, but also the associated comments section. Positioned in close proximity to the video, this section allows registered users to post messages and unregistered users to read them while viewing the content (Byun et al., 2023). Creators leave comments on their videos to receive feedback on content quality and viewer preferences. They also strategically respond to others' comments to encourage viewer interactions (such as likes, comments, shares, and subscriptions), thereby increasing the chances of YouTube's algorithm recommending their channel to other users (Hokka, 2021; YouTube, 2025). For viewers, the comments section is an integral part of the entertainment, networking and socialization experience, providing a space for communication with both creators and fellow audience members (Byun et al., 2023; Rotman and Preece, 2010).

However, while the comments section promotes expression and connectivity, some users exploit this space to discriminate against others and disseminate racist narratives. Sports events, particularly international ones such as the Euro championships and the Olympics, serve as focal points for expressions of racism and xenophobia (Kearns et al., 2023). The competitive nature of these events and the use of national symbols such as national flags and anthems tend to promote negative 'us versus them' dynamics (Cho, 2009), often involving fans' intense emotional reactions to athletes, teams, sports managers, coaches, referees, and other fans (Cleland, 2014; Miranda et al., 2023; Sanderson, 2013; Wann et al., 2005). The immediacy, anonymity, and accessibility of YouTube's comments sections amplify the hateful tone in ways that are less prevalent in conventional sports media (Litchfield et al., 2018: 155). Although YouTube channel owners can moderate comments by removing inappropriate content or blocking users, previous studies have indicated that many hesitate to do so, as they are aware that provocative comments often drive engagement and contribute to channel growth (Larsson, 2017; Madden et al., 2013; Matamoros Fernandez, 2018). As a result of this phenomenon, YouTube has become a source of data for research on how racism manifests in digital spaces, particularly within the emotionally charged context of international sports events (Thelwall, 2018).

Studying online racism within sports contexts has important implications. At the individual level, victims of online hate speech often experience psychological distress, including depression, distrust, and reduced sports performance (Bilewicz and Soral, 2020; Frischlich et al., 2021; Lishman et al., 2024; Näsi et al., 2015; Ștefăniță and Buf, 2021; Soral et al., 2018). For example, after the Euro 2020 tournament, sportspeople and policymakers spoke out about the emotional toll of the abuse directed at players, highlighting the need for greater support systems for teams and more robust measures to combat online racism (Back and Mills, 2021; MacInnes and Duncan, 2021; McVeigh and Hall, 2021). At the societal level, online racism has the potential to erode diversity in sports participation and viewership by shaping perceptions of who legitimately belongs within particular sports (Kavanagh et al., 2019; Kearns et al., 2023; Litchfield et al., 2018). Research has also shown that regular exposure to hateful social media narratives can prompt dominant-group members to socially and physically distance themselves from stigmatized minorities, reducing cross-cultural understanding (Gelber and McNamara, 2016; Leets, 2002; Meriläinen, 2024; Soral et al., 2020; Wilson, 2022). Moreover, the systematic and intense abuse of minority athletes can lead to a displacement of hate, which affects minority groups sharing the same identity as the stigmatized athlete (Nägel et al., 2024). Detecting hateful content and identifying vulnerable groups in the sporting ecosystem is, therefore, a key process in fostering safer digital and social environments and equality in sports (Ashraf et al., 2021).

One effective strategy for detecting racism on social media is the use of natural language processing (NLP) models. NLP techniques, such as word embedding and topic modeling, allow researchers to systematically collect and process large volumes of unstructured textual data. These methods help identify the frequency, targets, and linguistic strategies employed in racist narratives that would either otherwise remain hidden in plain sight, or require labor-intensive manual efforts (Mullah et al., 2024). Studies applying NLP to understanding racism in sports initially emerged from the fields of computer science and software engineering. These studies primarily used social media data to develop automated hate speech detection techniques and compare the performance of various models (for example, see Aljarah et al., 2020; Aloufi and El Saddik, 2018; Karayiğit et al., 2022; Pookpanich and Siriborvornratanakul, 2024; Vujičić Stanković and Mladenović, 2023). While there is a growing number of social science studies discussing the patterns of discriminatory online communication in sports, many of these continue to rely on human coding and qualitative analysis (for example, see Black et al., 2023; Glynn and Brown, 2022; Sanderson, 2013; Sanderson et al., 2016; Seijbel et al., 2022). Furthermore, a literature review by Kearns et al. (2023) reported that the focus of the existing research was almost exclusively on Twitter and Facebook, indicating the need for broader research on other digital platforms (for example, see Avcı and Kılınçarslan, 2023; Lingiardi et al., 2019; Seijbel et al., 2023; Miranda et al., 2023; Rudwick and Schmiedl, 2023). Kearns et al. (2023) also reported a limited geographical scope and range of sports disciplines in existing studies. Most analyses have focused on examining racist discourse in the United Kingdom and the United States, primarily in football/soccer and American football, reflecting their popularity and cultural significance (for example, see Fenton et al., 2024; Kilvington et al., 2022; Mullah et al., 2024; Poulton and Durell, 2016). Additionally, much of the research has focused on racism directed at prominent athletes, overlooking other possible sports figures who are potential targets of exclusionary boundary-making (for example, see Alsagheer et al., 2022; Koenigstorfer et al., 2023; Litchfield et al., 2018; Oshiro et al., 2020).

1.3 What do we do?

Using NLP models, this report expands the research on online racism in sports by examining YouTube comments in the context of four major international sports events: Euro 2020 and 2024, UEFA Women's Euro 2022, and the Paris Olympics 2024. Examining the Euro football championships allows us to explore how racism manifests within one of the most popular sports in the world. The UEFA Euro 2020 offers a unique case, with games played across 11 cities in different countries to celebrate the championship's 60th anniversary. It is likely that this multinational hosting approach amplified cross-cultural interactions, intergroup rivalries, and possibilities for boundary-making practices. Also, the Euro 2020 tournament was postponed to 2021 due to the COVID-19 pandemic and subject to various restrictions, including limited venue capacity, social distancing, mandatory face masks, the requirement for a negative COVID-19 test result, proof of vaccination, or proof of immunity for entry into stadiums, restaurants, and pubs for viewing games (BBC, 2021; Jahns, 2021). These restrictions may have increased reliance on digital platforms like YouTube for viewing games and expressing emotions and opinions, thus potentially increasing online racist narratives.

The inclusion of UEFA Women's Euro 2022 broadens the analytical scope to discrimination in women's football, providing a comparative perspective on how racist discourse may differ across gendered sports spaces. The addition of Euro 2024 also allows for a comparison of how the changes in social and political contexts between 2020 and 2024 influenced the dynamics of racist narratives. This period was notably characterized by shifts in athlete activism. For example, during Euro 2020, some players took the knee before matches to raise awareness about racism in football. However, in the 2024 tournament, there was more resistance to this practice, with some fans and commentators portraying players as 'puppets', and framing kneeling as politicization of the game (Dixon et al., 2023; O'Neill et al., 2025). The narratives of inclusion, exclusion, and xenophobia in sport may have been also affected by the Russian invasion of Ukraine in 2022 and the exclusion of Russian teams from the Euro games. The 2024 Olympics further enhances this work package's analytical reach with its multi-sport context. As the most international and culturally diverse sporting event, the Olympics are characterized by heightened national pride and global attention (Garcia, 2008). The 2024 Games were also distinctive due to the participation of Russian athletes under the status of "Individual Neutral Athlete", making it an interesting case for examining racialized and ethno-national forms of boundary-making.

Using a combination of NLP techniques, including machine translation, hate speech detection, topic modeling, and named entity recognition, this work package investigated how racism and xenophobia are expressed, targeted, and circulated in YouTube comments during these major international sporting events. Using this approach, it was possible to systematically examine the frequency and content of negative online interaction, identify the social identities and individuals most frequently targeted, and evaluate how sports organizations respond to such discourse in their YouTube channels. Rather than being a static or singular process, boundary-making is an everyday practice that is continually negotiated and reconfigured in response to shifting social, political, and cultural contexts. As Romero (2018) emphasized, boundaries are rarely constructed around a single axis of identity; rather, they emerge at the intersections of race, ethnicity, nationality, religion, gender, and sexuality. Online hate speech, particularly in the emotionally charged domain of international sports, often reflects these entangled dynamics. Racist discourse tends to reinforce wider systems of exclusion and domination, frequently co-occurring with sexism, homophobia, transphobia, nationalism, and religious hatred (Kavanagh et al., 2019; Litchfield et al., 2018). Our research efforts, therefore, aim not only to understand racism but also hate directed at gender/sexual minorities, nationalities, and religion.

Our methodological approach involved seven main steps. In the first, we designed search strings to retrieve YouTube content related to the four international sporting events, focusing on videos published during each event, as well as one month before and after, using the YouTube Application Programming Interface (API). Specifically, we conducted three separate searches. The initial search retrieved videos without applying geographic filters, enabling the collection of global content. The second search applied geographic filters to identify content popular in Belgium, Germany, Hungary, the Netherlands, Poland, and Türkiye. The final search specifically focused on the official YouTube channels of national football federations, Olympic committees, and national sports broadcasting agencies.

In the second step, we extracted publicly available data associated with each video retrieved in the search process using the YouTube Data API. This process resulted in the creation of three datasets: the Event Dataset (globally relevant videos), the Country Dataset (country-specific popular videos), and the Organizational Dataset (official videos). The third step involved the detection of the language of each comment using the Langdetect Python library. To enable cross-linguistic analysis across approximately 60 languages, all non-English comments were translated into English using the Google Translate API. In the fourth step, we employed TweetNLP (Camacho-Collados et al., 2022), a Python library designed for social media text analysis, to identify comments containing hate and offensive language related to race, ethnicity, religion, or gender. Each comment was labeled as either hateful or not. The fifth step involved conducting a descriptive analysis of the dataset. This included generating word cloud visualizations to explore the most frequently used terms in hate comments, providing an initial indication of lexical patterns and thematic distinctions. In the sixth, related words and comments were clustered into coherent topics using BERTtopic (Grootendorst, 2022) modeling and class-based Term Frequency-Inverse Document Frequency (c-TF-IDF). In the final step, each topic that emerged from the BERTopic analysis was examined to determine whether it included references to nationality or country, geographic locations, race or ethnicity, religious identity, gender, or specific individuals using Named Entity Recognition (NER). This process enabled us to identify the social identities and public figures most frequently targeted in the online hate discourse in the dataset.

2. Method

2.1 Dataset Construction

To retrieve relevant YouTube content related to four major international sporting events, we first identified appropriate search keywords. The search strings were developed based on commonly used terminology on the platform: Euro 2020, Euro 2022, Euro 2024, Paris2024, and Paris Olympics 2024. Each search string was applied within a standardized timeframe: the month of the event, as well as one month before and after. This three-month window was chosen to capture the peak period of public interest, including both pre-event anticipation and post-event reactions. Public interest typically increases around the time of a triggering event and diminishes rapidly in the aftermath (Burnap and Williams, 2015).

We then conducted three distinct sets of searches using the YouTube Data API, resulting in the construction of three separate datasets. The first used the keywords listed above without any country-specific filters, yielding 'the Event Dataset', which aggregates globally relevant videos related to the four tournaments. The second search applied geographic filters by specifying the *regionCode* parameter in the YouTube API. The following regional codes were used to create 'the Country Dataset': 'DE' for Germany, 'BE' for Belgium, 'HU' for Hungary, 'NL' for Netherlands, 'PL' for Poland, and 'TR' for Türkiye. This dataset captures popular content in these specific countries that attracted high levels of user interaction, such as views, likes, and comments. The third search retrieved videos from the official YouTube accounts of national football federations, Olympic committees, and public sports broadcasters in Belgium, Germany, Hungary, the Netherlands, Poland, and Türkiye, listed as in Table 1. This resulted in the 'the Organization Dataset.' National football federations and Olympic committees were chosen for three reasons: their authoritative roles in sports governance, their duty of care toward athletes, and their potential influence over public discourse (Kilvington and Price, 2019). However, in cases where these organizations either disabled comments or failed to provide event-specific content, data was instead collected from national sports broadcasting agencies. For example, because the official YouTube channel of Türkiye's football federation (@TürkiyeFutbolFederasyonu) had disabled comments on all videos, we instead used content from the national state-owned sports broadcasting channel, TRT Spor, for Euro-related events. For the Paris Olympics, national Olympic committees had posted either a very limited number or no videos, hampering our cross-organizational analysis, therefore we also included videos from national sports broadcasting agencies for all six countries.

Table 1. Official YouTube channels used for data collection: National sports organizations and broadcasting agencies by country

<table>
<tr><th>Event</th><th></th><th>Organization</th></tr>
<tr><td rowspan="6">Euro Games</td><td>Belgium</td><td>Royal Belgian Football Association (@royalbelgianfa)</td></tr>
<tr><td>Germany</td><td>German Football Association (@GermanFootball)</td></tr>
<tr><td>Hungary</td><td>Hungarian Football Federation (@MLSZTV)</td></tr>
<tr><td>Netherlands</td><td>Royal Dutch Football Association (@OnsOranje)</td></tr>
<tr><td>Poland</td><td>Polish National Football Association (@LaczyNasPilka)</td></tr>
<tr><td>Türkiye</td><td>TRT Spor (National Public Broadcaster Sports Channel, @trtspor)</td></tr>
<tr><td rowspan="13">Paris Olympics</td><td rowspan="3">Belgium</td><td>Belgian Olympic and Interfederal Committee (@BelgiumOlympic)</td></tr>
<tr><td>Sporza (Sports Service of the Flemish Public Broadcaster VRT, @Sporza_be)</td></tr>
<tr><td>RTBFSport (Radio-télévision belge de la communauté française, @RTBFSport)</td></tr>
<tr><td rowspan="2">Germany</td><td>German Olympic Sports Confederation (@DOSBvideo)</td></tr>
<tr><td>ZDF Sportstudio (German Public Broadcaster ZDF Sports Program, @sportstudio)</td></tr>
<tr><td rowspan="2">Hungary</td><td>Hungarian Olympic Committee (@olympicteamhungary)</td></tr>
<tr><td>M4 Sport (Hungarian Public Television Sports Channel, @M4Sport-Hivatalos)</td></tr>
<tr><td rowspan="2">Netherlands</td><td>Dutch Olympic Committee (@nocnsf)</td></tr>
<tr><td>Sports Service of the Dutch Public Broadcaster NOS (@nossport)</td></tr>
<tr><td rowspan="2">Poland</td><td>Polish Olympic Committee (@pkolpl)</td></tr>
<tr><td>TVP Sport (Polish Public Television Sports Channel) (@tvp_sport)</td></tr>
<tr><td rowspan="2">Türkiye</td><td>TRT Spor (Türkiye's National Public Broadcaster Sports Channel, @trtspor)</td></tr>
<tr><td>Turkish National Olympic Committee (@TürkiyeMilliOlimpiyatKomitesi)</td></tr>
</table>

From each video identified through these searches, all available comments were extracted using the YouTube API. For each comment, we retrieved the following metadata:

1. **Comment ID:** A unique identifier for each comment.
2. **Video Number:** A unique identifier for each video.
3. **Video URL:** The direct link to the video.
4. **Channel Name:** The name of the YouTube channel.
5. **Published At:** The publication date of the video.
6. **Video Likes:** The number of likes the video received.
7. **Keyword:** The search term used to retrieve the video (e.g., "Euro 2020," "Paris2024").
8. **Video_Comment ID:** The rank of the comment within the comment section.
9. **Author:** The anonymized username of the comment author.

10. **Comment Text:** The content of the comment.
11. **Comment Likes:** The number of likes the comment received.

2.2 Data Analysis

NLP, a subfield of artificial intelligence at the intersection of computer science, linguistics, and cognitive science, offers an effective and systematic approach for analyzing large-scale and complex social media content. Among the most widely used techniques in NLP are text summarization, machine translation, sentiment analysis, information retrieval, and speech recognition (Nadkarni et al., 2011). These tools enable researchers to classify textual data and uncover patterns in language. In this work package, NLP was applied to systematically analyze how racism and other forms of discrimination were expressed, targeted, and circulated in online conversations about sports. Specifically, the core methods we employed include word frequency analysis, NER, machine translation, hate speech detection, and topic modeling.

After extracting comments from the three searches, which formed the Event Dataset, the Country Dataset, and the Organization Dataset, we applied language identification to each comment during preprocessing. The distribution of languages for the Event Dataset is provided in Appendix 1. Due to the size of the tables, this report omits the language distributions for the Country Dataset and the Organization Dataset, but these data are available upon request. The datasets included comments written in 56 different languages, reflecting the linguistic diversity of YouTube users interested in major international tournaments. English was the dominant language across all events, comprising between 46% and 50% of all comments, followed by Italian (it), Turkish (tr), Polish (pl), Indonesian (id), Spanish (es), French (fr), Portuguese (pt), German (de), Somali (so), and Welsh (cy). We employed the Google Translate API to translate all non-English comments into English, to ensure linguistic consistency across the data. Machine translation refers to the automatic process of converting text from one language into another while aiming to preserve the meaning and nuances of the source language (Wang et al., 2022). This approach was essential for preparing the data for subsequent analysis, as most NLP models and tools are primarily trained on English-language corpora and is especially critical for low-resource languages like Somali and Welsh, with limited annotated datasets and open-source libraries.

Following this stage, to identify racist comments we performed hate speech detection, a natural language processing technique. This technique is more sophisticated than conventional sentiment analysis, which in most cases, simply classifies text as positive, negative, or neutral (Boiy et al. 2007). Instead, hate speech detection aims to detect aggressive, toxic, or insulting language directed at individuals or communities based on their social identity. This technique has become increasingly common in areas such as social media monitoring and online content moderation (Alkomah and Ma, 2022; Jahan and Oussalah, 2023). In this study, we employed the TweetNLP Hate Speech Detection library, which is a Python-based tool specifically designed to identify hatred and discrimination based on race, ethnicity, nationality, religion, gender, and sexual orientation (Camacho-Collados et al., 2022; Basile et al., 2019). Using linguistic and semantic indicators, TweetNLP enabled us to automatically classify each comment in our datasets according to whether or not it contained hate speech. To complement this analysis, we also employed the Offensive Language Identification feature of TweetNLP, which detects comments that contain "any form of non-acceptable language (profanity) or a targeted offense, which can be veiled or direct" (Zampieri et al., 2019: 75). Using this tool, we classified comments as either offensive or not. Finally, we combined those comments labeled as hateful and those labeled as offensive to create a new subset, referred to as the Hate Dataset.

We then conducted descriptive analyses of this dataset to better understand the distribution of the content. Specifically, we examined the word frequency patterns, the temporal distribution of comments, and the number of likes for hateful comments. Word frequency analysis was used to identify the most common words within this subset, offering insight into dominant themes and recurring language patterns. To enhance interpretability, we visualized these patterns through word clouds. As is common practice in the literature, removing stop words such as "the," "is," and "and" allowed us to focus on content-specific vocabulary with greater semantic value (Kwon et al., 2021; Turki and Roy, 2022; Xiang et al., 2021).

Then, we applied topic modeling to the subset of comments in the Hate Dataset. This unsupervised machine learning technique organizes textual data into distinct topics based on patterns of word co-occurrence. It is particularly useful in exploratory data analysis, enabling the identification of underlying themes without requiring manual coding or predefined categories (Abdelrazek et al., 2023; Kwon et al., 2021; Mujahid et al., 2021). In this work package, we employed BERTopic which combines transformer-based language models with c-TF-IDF to generate semantically meaningful topic clusters. A major advantage of BERTopic over more traditional techniques, such as Latent Dirichlet Allocation (LDA), is its effectiveness with short, informal texts such as YouTube comments. Using BERTopic, we clustered each dataset's hate-offensive subset into ten topic groups and examined the representative keywords within each in order to identify the most salient identity-related discourses.

To further explore the targets of hate speech, we applied NER, an NLP technique used to extract names of individuals from unstructured data (Ehrmann et al., 2023; Hu et al., 2024; Jehangir et al., 2023). In this work package, we utilized spaCy, a robust and efficient open-source NLP library in Python. Specifically, we employed spaCy's en_core_web_sm model, an English pipeline equipped with vocabulary, syntactic parsing, and NER components (Honnibal and Montani, 2017). This helped us automatically identify individual names in the BERTopic-generated keywords. Entities labeled as ‘person’ were then manually reviewed by the two senior researchers in this work package and inductively categorized into the following seven identity groups: athlete, political leader, television broadcaster/journalist, musician, referee, coach, and miscellaneous. To identify specific social-identity categories in the BERTopic clusters, we followed two steps. First, after running BERTopic on the hate-labelled comments in each dataset, the ten largest clusters per corpus and their top keywords were extracted (as ranked by cTF-IDF). The same two senior researchers independently examined the top cTF-IDF keywords for every cluster and flagged and recorded any token or phrase that functioned as a social identifier (whether a nationality, city, racial label, religious term, or gendered reference). Once all clusters had been reviewed, recurring words collected under five broad headings: Nationality/country, geography (place-based labels such as “Paris” or “Europe”), race/ethnicity, religious identity, and gender identity. No additional category appeared often enough to warrant a separate label.

3. Results

3.1 Distribution of videos and comments

Table 2 presents the number of videos retrieved, the total number of comments, and the average word count per comment in the Event Dataset. Euro 2020 stood out with the highest number of comments (279,238) and videos (565), possibly due to the tournament’s multi-country hosting format and the exceptional circumstances of the COVID-19 pandemic. Since matches were played across 11 different countries, attendance in stadiums was likely dominated by local audiences. Many international fans, unable to travel between geographically distant cities such as Baku, Rome, Budapest, and Glasgow, may have instead followed the games online, contributing to both the creation and demand for more video content. Additionally, COVID-19 restrictions still in place in 2021 may have led many to watch the matches remotely, turning to digital platforms to express their thoughts and emotions. The tournament’s timing might also have affected the video and comment counts. Euro 2020, the earliest tournament in the dataset, was not held until 2021 due to COVID-19 postponement, and thus had more time to accumulate views and comments.

However, this last explanation does not apply to Women’s Euro 2022. Despite being chronologically the second event in the dataset, it recorded the lowest number of comments (17,220) and videos (396), indicating lower levels of interest compared to men’s tournaments. Nevertheless, Women’s Euro 2022 recorded the highest average comment length (16.73 words), suggesting more elaborate user input. In contrast, Euro 2024 had the lowest average word count (9.01), which suggests more reactive comments during games. The Paris Olympics had fewer videos compared to the men's tournaments, but surpassed the Women's Euro 2022 in overall comment volume.

Table 2. Event Dataset: Video count, comment volume, and comment length by event

Event	Videos (n)	Comments (n)	Average comment length in words
Euro 2020	565	279,238	12.21
Women's Euro 2022	396	17,220	16.73
Euro 2024	482	183,710	9.01
Paris Olympics (keyword - Paris Olympics 2024)	198	36,957	15.05
Paris Olympics (keyword - #Paris2024)	246	154,219	11.0

Table 3 shows the distribution of YouTube videos and comments retrieved for the Country Dataset. Similar to the Event Dataset, Euro 2020 generated the highest level of interaction across all six countries. Popular videos in Türkiye (275,798 comments), Hungary (259,307 comments), and Poland (256,920 comments) had the highest levels of engagement. The average comment length was consistent across countries, between 11 and 12 words. For Euro 2024, compared to Euro 2020, in most countries there was a decline in both the number of comments and the average word count per comment. Popular videos in Türkiye once again led in both the number of videos and total comments in 2024. Women's Euro 2022 attracted fewer but longer comments, averaging over 16 words in most videos. Lastly, the Paris Olympics 2024 had the highest number of videos across all six countries. Despite more moderate comment volumes, ranging from 145,883 to 186,048, the average length was consistently higher than in Euro 2024, at around 11 words.

Table 3. Country Dataset: Video count, comment volume, and comment length by event

Event	Country	Videos (n)	Comments (n)	Average comment length in words
Euro 2020	**BE**	89	169,908	11.00
	DE	400	235,680	11.65
	HU	400	259,307	11.93
	NL	387	244,980	12.04
	PL	377	256,920	11.87
	TR	410	275,798	11.94
Women's Euro 2022	**BE**	418	16,478	16.69
	DE	429	16,415	16.66
	HU	393	14,990	16.48
	NL	426	17,063	17.06
	PL	432	18,165	17.18
	TR	417	16,520	16.67
Euro 2024	**BE**	351	200,463	7.38
	DE	337	215,009	9.06
	HU	342	196,172	7.35
	NL	125	131,195	6.61
	PL	358	207,067	7.25
	TR	365	257,028	7.76
Paris Olympics (keyword - #Paris2024)	**BE**	595	186,048	11.27
	DE	536	145,883	11.08
	HU	578	173,056	11.26
	NL	571	177,698	11.15
	PL	574	167,919	11.18
	TR	595	177,666	11.07

Table 4 presents the Organizational Dataset, summarizing the number of videos shared by national football federations, Olympic committees, and national sports broadcasters across six countries, along with the number comments received and the average comment length. Organizational activity on YouTube during major sports events appears uneven across countries. For both Euro 2020 and 2024, Türkiye's state-owned broadcaster TRT Spor posted more videos and attracted more viewer comments than other nation's official channels. For the Paris Olympics, Germany's Sportstudio led with 115 videos and 20,268 comments. Other countries' national channels had either no event related videos or posted fewer videos and received less public engagement. The Women's Euro attracted the least attention. This may be partly explained by participation, because only Belgium, Germany, the Netherlands, and Poland qualified. However, neither Germany and nor Poland's official channels posted any relevant videos. This pattern suggests that national federations and broadcasters may still prioritize men's football in terms of media visibility, resource allocation, and promotional strategy.

Table 4. Organizational Dataset: Video count, comment volume, and comment length by event

Event	Country	Channel name	Videos (n)	Comments (n)	Average comment length in words
Euro 2020	**BE**	royalbelgianfa	60	914	9.68
	DE	GermanFootball	-	-	-
	HU	MLSZTV	1	2	13.5
	NL	OnsOranje	-	-	-
	PL	LaczyNasPilka	3	2,318	17.36
	TR	trtspor	63	25,111	11.8
Women's Euro 2022	**BE**	royalbelgianfa	17	76	9.83
	DE	GermanFootball	-	-	-
	HU	MLSZTV	-	-	-
	NL	OnsOranje	3	80	15.8
	PL	LaczyNasPilka	-	-	-
	TR	trtspor	-	-	-
Euro 2024	**BE**	royalbelgianfa	37	1,321	12.27
	DE	GermanFootball	5	128	13.36
	HU	MLSZTV	-	-	-
	NL	OnsOranje	1	93	10.74
	PL	LaczyNasPilka	10	7,003	17.15
	TR	trtspor	111	61,634	13.39
Paris Olympics	**BE**	BelgiumOlympic	-	-	-
		Sporza_be	-	-	-
		RTBFSport	-	-	-
	DE	DOSBvideo	-	-	-
		sportsstudio	115	20,268	18.91
	HU	olympicteamhungary	2	5	9.4
		M4Sport-Hivatalo	-	-	-
	NL	nocnsf	-	-	-
		nossport	1	16	20.88

	PL	pkolpl	35	172	9.74
		tvp_sport	1	25	17.32
	TR	trtspor	6	389	9.34
		TürkiyeMilliOlimpiyatKomitesi	2	67	14.16

3.2 Distribution of hate and non-hate comments

Table 5 displays the distribution of comments classified as hate or non-hate in the Event Dataset. To facilitate interpretation, the proportion of hate comments has been emphasized with a color scale, with darker colors representing higher values. Overall, 10.1 percent of all comments were identified as hateful or offensive. Among the four events, the Paris Olympics 2024 stood out with the highest proportion of hate-labeled comments, at almost 19 percent. This finding suggests that the multi-sport and global nature of the Olympics may intensify national rivalries and other in-group/out-group tensions. One group of earlier studies have argued that mega-events like the Olympics can "have socially integrative effects by bringing together a diverse audience engaged and immersed in a common activity" (Kersting, 2007; Ismer et al., 2015: 554). In contrast, another group of studies has pointed to these events' "socially disintegrative effects", suggesting that they can also "promote exclusion and the rejection of minorities and prejudiced groups by forcing people to confront each other in competitive and affectively laden ways" (Ismer et al., 2015: 554; Rolf, 2023). Women's Euro 2022, despite having the fewest total comments overall (17,220), exhibited the second highest proportion of hate speech (around 10 percent), highlighting the persistent challenges facing women in online spaces (Fenton et al., 2024). By contrast, Euro 2024 had the lowest proportion of hateful comments, at around six percent, suggesting a less hostile online environment. In absolute terms, Euro 2020 generated the highest number of hate comments (21,972).

Table 5. Distribution of hate and non-hate comments in the Event Dataset

Event	Hate comments (n)	Non-hate comments (n)	Total comments (n)	Hate (%)	Non-Hate (%)
Euro 2020	21,972	257,266	279,238	7.9	92.1
Women's Euro 2022	1,701	15,519	17,220	9.9	90.1
Euro 2024	10,083	173,627	183,710	5.5	94.5
Paris Olympics 2024	6,835	30,122	36,957	18.5	81.5
#Paris2024	13,557	140,662	154,219	8.8	91.2
Total	54,148	480,430	534,578	10.1	89.9
Average	10,829.60	12,3439.20	-	8.1	91.9

Table 6 presents an overview of hate comments that received ‘likes’ across the four events. Likes serve two important roles on YouTube. First, they contribute to the platform’s recommendation algorithm, influencing which videos are promoted to wider audiences, and those with a higher number often gain greater visibility (Matamoros-Fernández et al., 2021). Second, likes serve as a form of social validation. YouTube does not display the number of ‘dislikes’ publicly, therefore, likes have become the primary metric of audience feedback on comments. The more likes, the greater the comment’s resonance with viewers (Scissors et al., 2016). In the case of hate speech, high counts of likes, without the counterbalance of visible disapproval, may create the illusion of widespread social approval, potentially contributing to the normalization of harmful views.

As shown in Table 6, Euro 2020 recorded the highest number of hate comments with likes (8,310 in total), accounting for 38 percent of all hate comments in that dataset, with an average of 25 likes per hate comment. This tournament’s multinational hosting format likely increased intergroup rivalries and possibilities for boundary-making. The international competition, along with “anxiety and distress triggered” by high stakes matches, may have exacerbated supporters’ perceptions of opposing teams, fans and diverse host communities as members of an out-group (de Leon and Kim, 2016: 9; Ismer et al., 2015). The COVID-19 pandemic may have further intensified existing social divisions, including polarization, racism, and discrimination. In many societies, at this time, the prevailing uncertainty, and fear, along with the rapid spread of misinformation created fertile ground for scapegoating and prejudice. Previous research, for example, has identified a sharp increase in conspiracy theories targeting Jews and Chinese communities during this period. These dynamics may have been reflected in the volume of online hate speech and the validation of such narratives through user engagement, such as ‘likes’ (Dai et al., 2024).

Table 6. Distribution of likes on hate comments in the Event Dataset

Event	Number of hate comments with likes	Total number of likes in hate comments	% of hate comments with likes	Average number of likes per comment
Euro 2020	8,310	211,579	38	25
Women's Euro 2022	746	3,341	44	4
Euro 2024	2,764	73,523	27	27
Paris Olympics 2024	1,758	116,887	26	66
#Paris2024	5,049	191,709	37	38
Average	3,725	119,408	34	32

Women's Euro 2022 had the highest percentage of hate comments with likes (44 percent); yet, it also had the lowest average number of likes per hate across all events. This pattern suggests that women's football remains a contested space, yet the relatively low number of likes per hate implies only limited endorsement of hateful views by other users. The comment that received the most likes (112 in total) was: "Love the fact that mum and dad could take the kids and enjoy a game without the hordes of drunken young men wanting to fight each other you get at the men's games." This comment reflects a gendered perception of football spaces, in which the women's game is framed as being more family friendly. The praise here is not for the quality of play or athleticism, but for the absence of aggressive masculinity. It is important to note, however, that like counts should be interpreted with caution because a small number of highly active users or even a single individual could have been responsible for liking multiple hate comments. Nonetheless, the presence of such engagement contributes to the normalization of discriminatory discourse.

Table 7 shows the distribution of hate and non-hate comments in the Country Dataset, which includes user comments retrieved from videos that were popular in Belgium, Germany, Hungary, the Netherlands, Poland, and Türkiye. The overall proportion of hate comments is relatively consistent across countries. The videos popular in the Netherlands attracted the highest proportion of hate comments at 7.4, slightly above the average. While the videos popular in Turkey had the largest absolute number of hate comments (48,404), this was proportional to their overall high volume of comments. The similarity in proportions of hate comments to total comments across countries suggests that hateful discourse was not confined to a specific national viewership.

Table 7. Distribution of hate and non-hate comments in the Country Dataset

Country	Hate comments (n)	Non-hate comments (n)	Total comments (n)	Hate (%)	Non-Hate (%)
BE	40,496	532,401	572,897	7.1	92.9
DE	41,138	571,849	612,987	6.7	93.3
HU	45,652	597,873	643,525	7.1	92.9
NL	42,154	528,782	570,936	7.4	92.6
PL	45,581	604,490	650,071	7	93
TR	48,404	678,608	727,012	6.7	93.3
TOTAL	263,425	3,514,003	3,777,428		
AVERAGE	43,904.20	585,667.20	-	7	93

Figure 1 shows the monthly distribution of comments on videos that were popular in six countries from May 2021 to September 2024. The highest peak across the entire timeline was in July 2021, coinciding with the quarter-finals (2–3 July), semi-finals (6–7 July), and final (11 July) of Euro 2020. Another major peak was in July 2024, aligning with the knockout phase of Euro 2024 tournament (5-14 July) and the start of the Paris Olympics (26 July). High levels of commenting continued into August 2024, at the final stages of the Olympic Games. In contrast, July 2022, which corresponds to the Women's Euro 2022 (6–31 July), showed a significantly lower volume of comments across all six countries. Similarly, comment volumes were consistently low in June 2022, May 2024, and September 2024, which coincided with periods without major international sporting events. This pattern is consistent with previous literature on hate speech, revealing that high profile events such as sports tournaments, elections, and political rallies receive broad media coverage, trigger emotional responses, activate collective identities, and become more susceptible to boundary-making discourse (Hobbs et al., 2024). Public interest and emotional investment typically increases around high-stake moments, leading to more reactive and provocative behavior, but diminishes quickly in the aftermath (Burnap and Williams, 2015; Nägel et al., 2024).

Figure 1. Monthly distribution of hate comments on videos popular in the Country Dataset

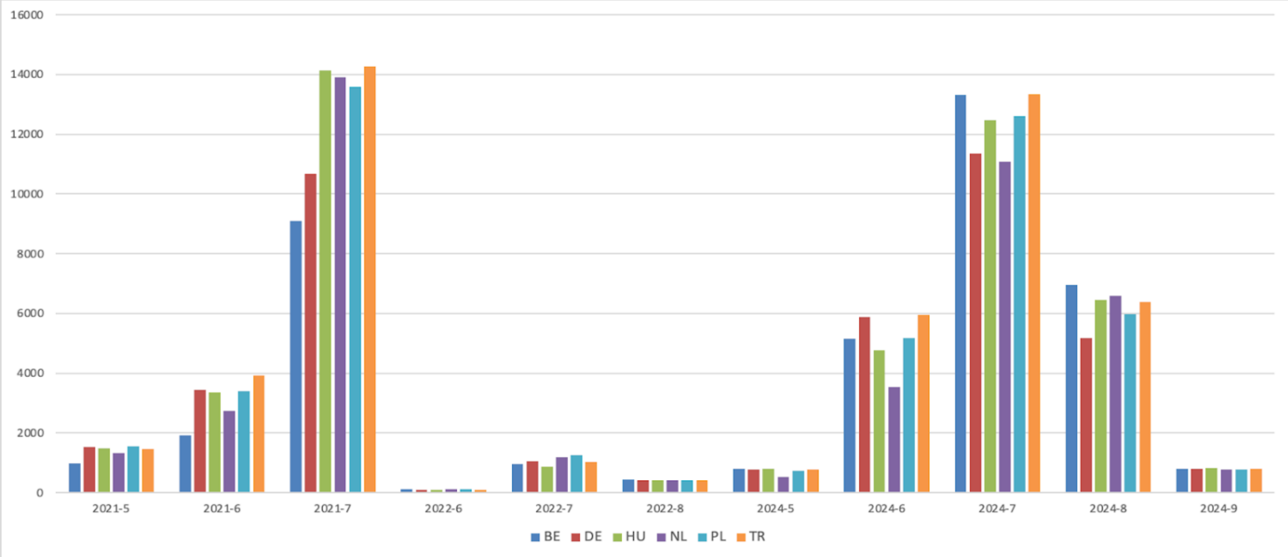

Table 8 shows the likes for comments posted under videos that were popular in six countries. On average, 34 percent of comments across the six countries received at least one like. The average number of likes per comment was 30. Videos popular in Belgium had the highest average number of likes per comment (36), despite having the lowest percentage of comments receiving likes (33 percent).

Table 8. Distribution of liked hate comments in the Country Dataset

Country	Number of comments with likes	Total number of likes	% of comments with likes	Average number of likes per comment
BE	13,493	481,964	33	36
DE	14,315	455,679	35	32
HU	15,917	425,418	35	27
NL	14,491	470,761	34	32
PL	15,748	438,672	35	28
TR	16,553	431,726	34	26
Average	15,086	450,703	34	30

3.3 Who is targeted? Individual and identity-based discrimination

Two analyses were performed on YouTube comments classified either as hate or non-hate in the Hate Dataset: the first focused on discrimination targeting specific individuals, and the second, on patterns of identity-based discrimination.

3.3.1 Individual targets of online hate

Table 9 presents the proportion of hate comments targeting individuals across the four sport events. We grouped targeted individuals by occupation: athletes, referees, coaches, television broadcasters/journalists, political leaders, musicians, and miscellaneous. Each cell shows the percentage of comments that reference a specific group within the hate dataset for the corresponding tournament. Values listed as <.001 indicate that hate comments targeting the group were at a minimal level. The heatmap shading shows the intensity of mentions within each event.

Table 9. Distribution of hate comments targeting individuals by event (%)

Category	Euro 2020	Women's Euro 2022	Euro 2024	Paris 2024	Paris Olympics 2024	Average
Athlete	1.584	.47	5.742	.745	.454	1.799
Referee	<.001	.235	<.001	<.001	<.001	.047
Coach	.687	.294	1.18	<.001	<.001	.432
Political leaders	.064	<.001	<.001	.288	.717	.214
Television Broadcaster/Journalist	<.001	.529	<.001	<.001	<.001	.106
Musician	.046	<.001	<.001	.369	.19	.121
Miscellaneous	.155	.176	.06	<.001	.088	.096

Across all tournaments, the highest concentration of targeting of named individuals occurred during Euro2024, at almost seven percent of all hateful comments, followed by Euro 2020, with slightly more than 2.5 percent of hate comments involving a named individual. Videos about the Paris Olympics featured the lowest levels of individual targeting. Women's Euro 2022, though marked similarly by a relatively lower share of personal hate (under two percent), showed a wider range targeted, including athletes, referees, coaches, and television broadcasters/journalists.

Athletes were the most frequently targeted in every event, especially in Euro 2024 and Euro 2020, in line with previous studies showing that this group is common targets of online abuse, hate against them often reflected "the ongoing construction of national identity and symbolic borders" (Kearns et al., 2023; Leung et al., 2024: 2). For example, comments like "[A1] the Nigerian ruined England national team" and "[A2] is brazilian, we're talking about euro, europe nations" illustrated how players with dual heritage were othered, even when representing national teams. Other remarks such as "[A3]. Like the rest of England is racist. Instead of being humble" or "many of the Italy players were dirty and disgusting" reflected blanket stereotyping of countries or teams, distinguishing the commentator's group from the 'other' and reinforcing in-group superiority. In some cases, hate speech intersected with race, sexuality, and nationalism, as seen in a comment like: "[A4]'s rainbow armband is currently wrapped around [M1]'s b*ll***k; apparently, England loves its gays but hates its blacks".[1]

Coaches were also frequently targeted, particularly in Euro 2024, Women's Euro 2022, and Euro2020. When teams underperformed, coaches were blamed for preferring Black or foreign-born players over White players. For example, one user stated: "[C1] made his own bed, and now he should lie in it, after his political diversity is our strength; power play back-fired... Totally Woke… Totally broke… Totally unbelievable." Such comments marginalized and delegitimized inclusionary tactics and multicultural values. Another user wrote: "bigger countries with bigger population like France, England, Germany etc. take black players in their NATIONAL TEAM. Even if the blacks are born there they should not be allowed in the national teams. Club level is fine to a degree but national level should be played by sons of soil as God, Universe and nature put them there. Genetic material and national pride is missing in black/Arab players when they play for white European countries." In this example, national teams were framed as sites for articulating exclusionary notions of belonging while portraying racial hierarchies as part of a divine and natural order, and Black and Arab players as incompatible.

Television broadcasters/journalists attracted the most attention in Women's Euro 2022. For example, a comment framed the promotion of women's football by the media as a "falsely egalitarian and feminazi injunction". This form of defensive othering sought to protect traditional gender roles and male dominance in sport. Political leaders were more visible in the Paris Olympics and Euro 2020, sometimes framed in dehumanizing and delegitimizing terms, such as "zombie", "demonic", "satanist" or "imperialistic gay". Other comments featured sexualized personal attacks to leaders, such as "[P1] essentially offered herself [...] to the referees." In such comments, commentators reaffirmed the superiority of their own identity (national, religious, and heterosexual) while positioning these political figures, and by extension, their nations, as corrupt and threatening outsiders. Historical grievances and colonial legacies were also invoked to justify hostility and social alienation, as in: "All the countries hate the UK because of

[1] All quoted YouTube comments are presented as they originally appeared, preserving the original spelling, grammar, punctuation, and use of cases. To protect privacy, personal names have been anonymized using the initial capital letter of the respective occupational category: for example, athlete names are replaced with A1, A2 and musicians as M1, M2, political leader as P1, P2, etc. Quotations from comments include examples of exploitative and offensive language, which has been partially masked using asterisks.

the British Empire" or, "stealing from African countries, giving nothing back." Musicians, referees, and miscellaneous figures were mentioned at lower frequencies across all events, however, the volume of comments about referees in Women's Euro 2022, and the miscellaneous category in Euro 2020, showed that viewer hate can extend beyond high-profile figures such as athletes and coaches to peripheral individuals within the sporting ecosystem.

Table 10 presents the distribution of hate comments targeting specific individuals in videos that were popular in each of the six countries analyzed. Each cell represents the percentage of comments referencing individuals in the hate dataset. Values listed as <.001 indicate that the group was targeted at a very low frequency. The color scale shows the intensity of targeting by country. As in Table 9, athletes were the most frequently targeted individuals in hate-labeled comments across all six countries, with figures ranging from 2.099 percent (Netherlands) to 2.334 percent (Germany). On average, slightly over 2.2 percent of hate comments referred to an athlete. Coaches ranked second, with values ranging from .222 percent (Belgium) to .382 percent (Netherlands). This finding aligns with previous research showing that individuals responsible for performance including "coaches, parents, administrators and athletes" are "potential victims" within the sports environment (Stirling 2009: 1091).

Musicians, though not central actors in the events, also appeared across all countries, potentially due to performances during opening and/or closing ceremonies, while less frequently targeted were politicians, television broadcasters/journalists, referees and a miscellaneous group including celebrities and athletes' family members.

Table 10. Distribution of hate comments targeting individuals in the Country Dataset (%)

Category	BE	DE	HU	NL	PL	TR	Average
Athlete	2.185	2.334	2.21	2.099	2.198	2.279	2.218
Referees	.005	.005	.004	.005	.004	.004	.005
Coach	.222	.374	.361	.382	.309	.355	.334
Political leaders	.22	.168	.151	.173	.156	.095	.16
Broadcaster/journalist	.022	<.001	.02	<.001	.009	.008	.01
Musician	.254	.163	.212	.218	.193	.171	.202
Miscellaneous	.163	.16	.079	.135	.097	.06	.116

Figure 2 presents word clouds showing the most frequently used words in hate comments involving the names of athletes, political leaders, and coaches across two datasets: the Event Dataset (panel a) and the Country Dataset (panel b). Larger words appear more frequently in the comments. Similar themes emerged across these datasets. The most important finding is that nationalistic targeting was central in hateful commentary. In comments referencing athletes, the most frequent words were “England,” “penalty,” “Italy,” “win,” and “team”, likely related to the Euro 2020 final. This tournament marked a significant moment for the England men’s football team, reaching the final of a major international championship for the first time since the 1966 World Cup. However, following their defeat to Italy in a penalty shootout, three Black players who missed their kicks reportedly received racist abuse (Mauro, 2024). Our findings are consistent with previous research showing the online abuse these three players received (Nägel et al., 2024), as well as studies showing that minority players often become targets of online hate following high stakes matches (Arday, 2023).

Figure 2. The word clouds of hate comments targeting athletes, politicians and coaches in Event (a) and Country (b) data sets.

(a) (b)

The England team for Euro 2020 was also notable for its multicultural composition, “with fifteen members of the squad having dual or multiple nationality, represented a new, more inclusive understanding of English identity” (Bennett, 2025: 2). Research has shown that such diversity can make teams vulnerable to racialized and xenophobic commentary, particularly in the emotionally charged contexts of sports, and especially following disappointing performances (Cleland, 2014; Kearns et al., 2023). Therefore, in this context, hate comments targeting the England team are likely to reflect broader societal anxieties about group boundaries and boundaries of inclusion.

Lastly, in both datasets, hate comments targeting politicians had a broader international scope, frequently referencing countries such as the United States, India, and France. Drawing on social identity theory, previous research has explained identity insults as a strategy to enhance positive self-esteem. According to this perspective, “Group X creates an identity insult by attributing to Group Y negative features, evil motivations, or foul values or by accusing Group Y of performing destructive or erroneous actions” (Korostelina, 2014: 217). The frequent use of aggressive and insulting terms including “uncivilized”, “corrupt”, “evil”, “racist”, and “shame”, as shown in Figure 2, implies that these comments involved criticisms beyond the games, embracing broader forms of national, political and cultural antagonism.

Table 11 shows the number of hate comments targeting individuals in the Organizational Dataset. Compared to other videos analyzed above, there are relatively fewer comments, and, correspondingly fewer hate comments on videos shared by national football federations, Olympic committees, and public broadcasters. A total of 119,622 comments were collected from organizational channels, of which 5,510 (about 4.6%) were labeled as hate comments. The results revealed that the greatest number of hate comments (2,202) were posted on videos by Türkiye’s TRT Spor during Euro 2024, followed by Germany’s Sportstudio (1,678 hate comments) during the Paris Olympics 2024. During Euro 2020, the Polish (Laczy Nas Pilka) and Turkish (TRT Spor) channels again received a higher number of hate comments (187 and 921 respectively). In contrast, the Women’s Euro 2022 had very low overall comment counts and minimal hate comments, possibly due to lower visibility or moderated interactions on institutional channels. As above, the majority of hate comments (99) targeted athletes, followed by politicians (7), musicians (7) and coaches (7), while there were less frequent mentions of other sport and public figures such as journalists (1), referees (2), and miscellaneous figures (2). The results also revealed that none of the organizations examined here responded to hateful comments on YouTube.

The low volume of comments might be due to organizational channels’ lack of attractiveness compared to fan-led or entertainment-focused channels. Their content is often more formal, neutral and information-focused, such as press conferences, training clips, or official highlights, generating potentially less interest then viral or controversial videos. Moreover, official channels are more likely to apply content moderation, either manually or algorithmically, to maintain a professional image and prevent reputational damage. Hate comments may be deleted quickly or suspended pending review, leading to fewer visible offensive remarks. Viewers may also behave more cautiously when commenting on official platforms. The presence of institutional branding (such as federation logos or official broadcaster names) can create a sense of surveillance or professionalism, discouraging users from posting inflammatory content.

Table 11. Distribution of hate comments targeting individuals in the Organizational Dataset (n)

Event	Organization	Number of Comments	Number of Hate Comments	Athlete	Referee	Coach	Politician	Television Broadcaster/ Journalist	Musician	Miscellaneous
Euro 2020	LaczyNasPilka	2,318	187	2	0	0	0	0	0	0
Euro 2020	royalbelgianfa	914	34	0	0	0	0	0	0	0
Euro 2020	trt-spor	25,111	921	21	0	3	0	0	0	0
Euro 2020	MLSZTV	2	0	0	0	0	0	0	0	0
Euro 2022	OnsOranje	80	10	0	0	0	0	0	0	0
Euro 2022	royalbelgianfa	76	5	0	0	0	0	0	0	0
Euro 2024	GermanFootball	128	3	0	0	0	0	0	0	0
Euro 2024	LaczyNasPilka	7,003	368	3	0	0	0	0	0	1
Euro 2024	OnsOranje	93	4	0	0	0	0	0	0	0
Euro 2024	royalbelgianfa	1,321	76	0	0	0	0	0	0	0
Euro 2024	trt-spor	61,634	2,202	67	0	4	0	0	0	0
Paris Olympics 2024	nossport	16	1	0	0	0	0	0	0	0
Paris Olympics 2024	pkolpl	172	1	0	0	0	0	0	0	0
Paris Olympics 2024	sportstudio	20,268	1,678	6	2	0	7	1	7	1

Event	Organization	Number of Comments	Number of Hate Comments	Athlete	Referee	Coach	Politician	Television Broadcaster/ Journalist	Musician	Miscellaneous
Paris Olympics 2024	trtspor	389	14	0	0	0	0	0	0	0
Paris Olympics 2024	Türkiye Milli Olimpiyat Komitesi	67	3	0	0	0	0	0	0	0
Paris Olympics 2024	tvp_sport	25	3	0	0	0	0	0	0	0
Paris Olympics 2024	olympicteamhungary	5	0	0	0	0	0	0	0	0
Total		119,622	5,510	99	2	7	7	1	7	2

3.3.2 Identity-based discrimination

Table 12 reports the share of hate-labelled comments that explicitly mention one of six social-identity categories (nationality/country, geography, race/ethnicity, religious identity, and gender identity) across (i) four major tournaments and (ii) videos popular in six countries. Percentages indicate the proportion of all hate comments that contain at least one reference to one of the categories. The final column ("Total") shows the cumulative percentage of hate comments that contained any identity cue (some comments reference multiple categories, so totals exceed 100 percent in some cases). The heatmap shows the intensity of mentions within each event.

The results for Event Dataset revealed that references to nationality or country were the most common in hate comments across all events, with the exception of Women's Euro 2022, where gender references were more frequent. In the case of the Paris Olympics, which featured a more diverse set of athletes and teams compared to the European Championships, comments also included more frequent references to locations and religion. Turning to the Country Dataset, once again, nationality and gender emerged as the most frequently mentioned social identity categories in hate comments. This is unsurprising, as previous studies have shown that, in international sporting events, athletic achievements serve as a proxy for "national health and well-being", and sport performance becomes closely tied to collective identity and status (Vincent et al., 2010: 201). In games, national symbols, such as flags, anthems, team colors, also activate collective belonging (Marivoet, 2006). The prominence of nationality is further amplified by intense media coverage, which often frame international competitions in highly emotional and even militaristic terms, as ‘battles’ (Bertoli, 2017: 836; Rosenzweig and Zhou, 2021; Vincent et al., 2010). In hate comments on the women’s tournament, gender was the most frequently mentioned category, further supporting prior studies’ arguments that “women in football consistently struggle against sexism and gendered stereotypes to be recognized as legitimate participants and members of the football community” (Eime et al., 2022; Forbes et al., 2015: 523; Harris, 2001).

Table 12. Distribution of identity references in hate comments (%): Event vs. Country Datasets

		Nationality/ Country	Geography	Race/ Ethnicity	Religion	Gender	Total
Event Dataset	**Euro 2020**	41.2	1.9	1.8	0.2	8.0	53.1
	Women’s Euro 2022	21.6	0.9	4.8	0.2	44.9	72.4
	Euro 2024	22.8	0.5	0.5	0.4	13.4	37.7
	Paris Olympics 2024	18.5	5.3	1.3	9.5	19.8	54.5
	Paris2024	22.2	3.4	2.6	2.2	13.0	43.5
Country Dataset	**BE**	27.2	2.1	1.6	1.0	13.4	45.3
	DE	26.6	1.6	2.0	0.8	12.5	43.4
	HU	28.2	2.0	1.6	0.9	12.7	45.4
	NL	28.4	2.2	2.1	0.9	12.4	45.9
	PL	28.0	2.0	1.7	1.0	13.1	45.8
	TR	27.9	2.0	1.7	0.9	12.7	45.1

Table 13 presents the frequency of identity-related references in hate comments posted along with official organizations’ YouTube videos during the four major sports events. Due to the relatively low number of total comments in the Organizational Dataset, absolute figures but not percentages are reported to avoid overinterpretation. Across all events, the most frequently mentioned identity category was nationality/country, a trend especially pronounced during Euro 2024, on videos posted by TRT Spor (Türkiye) and Laczy Nas Pilka (Poland). This pattern is consistent with findings from other datasets. In contrast, there were fewer references to geographic locations, race/ethnicity and religion. Religion-related hate was most visible in comments under Sportstudio videos during the Paris Olympics 2024, with 28 comments flagged. Gender identity was mentioned 504 times in total, the highest being in Euro 2024, again under videos shared by TRT Spor and Sportstudio.

Table 13. Distribution of identity references in hate comments (n): Organizational Dataset

Event	Organization	Nationality/ Country	Geography	Race/ Ethnicity	Religion	Gender Identity
Euro 2020	LaczyNasPilka	9	1	0	0	7
Euro 2020	royalbelgianfa	4	0	0	0	5
Euro 2020	trt-spor	145	2	2	2	83
Euro 2020	MLSZTV	0	0	0	0	0
Euro 2022	OnsOranje	1	0	0	0	5
Euro 2022	royalbelgianfa	1	0	0	0	1
Euro 2024	GermanFootball	0	0	0	0	0
Euro 2024	LaczyNasPilka	28	0	0	1	24
Euro 2024	OnsOranje	1	0	0	0	1
Euro 2024	royalbelgianfa	11	0	0	0	7
Euro 2024	trt-spor	412	3	5	3	197
Paris Olympics 2024	nossport	0	0	0	0	0
Paris Olympics 2024	pkolpl	0	0	0	0	0
Paris Olympics 2024	sportstudio	291	14	1	28	170
Paris Olympics 2024	trtspor	3	0	0	0	1
Paris Olympics 2024	Türkiye Milli Olimpiyat Komitesi	2	0	0	0	0
Paris Olympics 2024	tvp_sport	0	0	0	0	3
Paris Olympics 2024	olympicteamhungary	0	0	0	0	0
Total		908	20	8	34	504

Table 14 presents the most frequent identity-related tokens identified by BERTopic within hate-labeled comments from the Event Dataset, grouped into five categories. Each token's weighted frequency (WF) is also reported. The national category is dominated by references to England and the English, followed by other high-profile contestants such as Italy, France, Spain, China, and Germany, reflecting how international tournaments re-activate national rivalries. Many of the related hate comments specifically targeted English fans for their alleged racism, hooliganism, or disrespectful behavior during matches. These comments also included a range of intergroup insults to create national stereotypes. Other comments attacked non-white or non-Christian players in European national teams and challenged the legitimacy of players not conforming to a narrow ethno-religious view of who should be eligible to take their place in a national team, and by extension, in the nation itself. For example, one comment, "Germany is an Islamic country lol?" used mockery to question the presence of Muslim players in the German national team. In another example, "France team = not one white man in there lol. Might as well be called Uganda or something", whiteness was treated as the standard of national belonging. Contrary to expectations, Russia and Ukraine received comparatively limited attention in the dataset. In the comments where Russia was mentioned, references often focused on its exclusion from international sports competitions. For example, one comment discussed this at length: "Yeah USA invides just because they wsnt to colonize some country and get their resources. Before this Isreal invasion on Palestine they invided Syria together with USA because they are muslims where muslims shouldnt be; then they attacked Palestine because Hamas kidnapped 400 people. Smart ass Ukrainan biggest military battalion AZOV killed 30 thousand Russian because they are Russians during six years before start of the conflict. Non off them are good, still that's politics non off them should be banned and especially not banning just one." Such comments invoke historical and geopolitical grievances and reflected a broader cynicism toward international politics.

Commenters referred not only to countries but also to symbolic or politically charged places (for example, Wuhan, Gaza, Tiananmen), often in relation to conspiracy theories regarding COVID-19 or historical grievances (for example, "PRC uses more advanced virusbased enhancements invisible to anti-dope kits. Thanks to the bat lady from Wuhan Institute of Virology"; "It appears the CCP didn;t take his wuhan bat potion!"). Hate comments also invoked a wide range of gender identifiers, which were overt indications of both sexism and transphobia. The frequent use of terms such as man, women, straight, trans, homosexual, and lgbt highlights how derogatory speech targets debates around inclusion, masculinity, and gender roles (for example, "Football is for men's, not for women's! What the hell is that? That shows only girls playing football??"; "Maybe England man team waiting for the death of queen to win the Euro or World Cup. It's shameful the country full of bright and smart people is led by a woman."; "Olympics are now the biggest trash. It's no longer sports, it's now political propaganda for LGBT."; "Olympics is about sports not about naked LGBTQIA+ people mocking Christianity.") Regarding religious identity**,** Christians and Christianity appeared most often, but there also were mentions of Islam, Muslim, Jews, doctrinal divisions (Catholics, Shiite) and other identifiers (satanists, disbelievers). Two example comments in this category are: "So called christians and christian enthusiasts are upset about the Olympics opening ceremony but are freaking silent about the pedophilia amongst christian priests and clergymen!!!.", and "Bruh this guy is Muslim he should be focusing on his religion and football not girls it is prohibited").

Table 14. Most occurring identity-related keywords

	Nationality / Country		Geography		Gender		Race/Ethnicity		Religion	
#	word	WF	word	WF	word	WF	word	WF	word	WF
1	england	23,355	europe	3,247	gay	6,792	black	2,659	christians	746
2	english	10,903	paris	1,978	man	6,678	white	1,829	christianity	647
3	italy	7,798	west	738	guy	5,200	african	483	christian	497
4	france	6,627	africa	373	women	3,740	asian	285	muslim	410
5	french	3,084	wuhan	63	guys	3,226	chinol	173	islam	273
6	spain	2,743	gaza	48	men	3,092	arab	117	satanists	231
7	china	2,679	vatican	25	girl	2,353	arabic	40	muslims	223
8	italian	2,588	tiananmen	12	woman	2,064	ching	36	satanist	136
9	germany	1,974	jasenovac	7	boy	963	nigger	14	jude	95
10	portugal	1,724	londoner	7	girls	837	laz	6	atheist	75
11	chinese	1,548			lady	782			catholics	50
12	turkey	1,490			female	562			jew	50
13	british	1,480			straight	556			scientology	30
14	white	1,434			lgbt	477			scientologist	18
15	usa	1,127			male	471			shiite	7
16	german	1,114			womens	374			disbelievers	1
17	taiwan	988			ladies	342				
18	denmark	968			trans	320				
19	india	915			lads	308				
20	russia	913			mens	244				
21					homosexual	180				

Figure 3 displays word clouds for the two most frequently referenced identities in hate-labelled comments: Nationality/country (upper sub-panels) and gender identity (lower sub-panels). Panel (a) draws on the Event Dataset, while panel (b) uses the Country Dataset. The larger the words, the more frequently they appear in the respective comment sets; offensive terms have been blacked out. The results show that the word "fans" was most frequently used in both datasets, indicating that hateful discourse is frequently aimed at supporter groups rather than the athletes alone. The gender-identity panels reproduce much of the same football lexicon, yet focus more centrally on football and evaluative verbs (e.g., like, think, see, know), suggesting that gender-based hostility is often expressed through opinionated, conversational language, rather than mere slurs.

Figure 3. The word clouds of hate comments referencing nationality / country and gender identity in Event (a) and Country (b) datasets

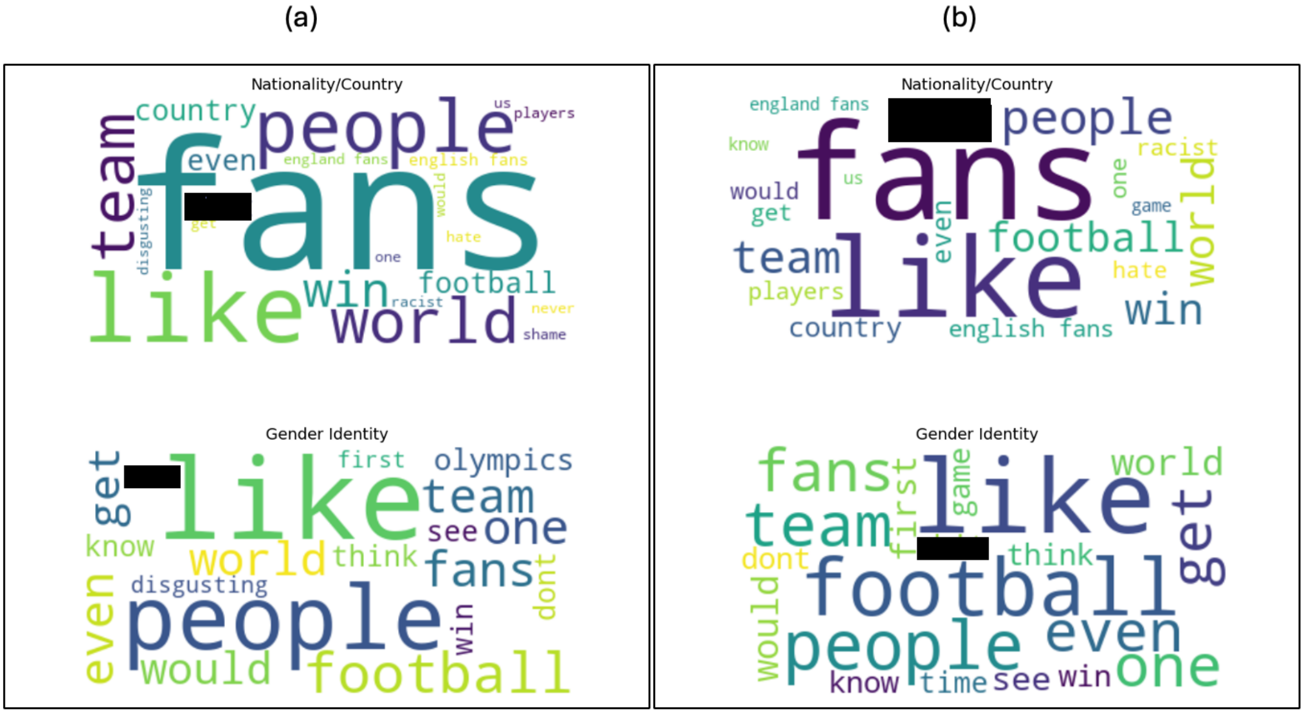

Table 15 presents the percentage of hate comments from the Event Dataset that include both an individual-level reference and an identity-related keyword. The largest number of hate comments referred to athletes and a nationality or country, and the second largest, to athletes and a gender identity. This supports the argument that athletes' performance and appearance have become sites for projecting racial, national, or gender-based anxieties. In terms of being targets of hateful comments, athletes were followed by coaches, particularly again with nationality and gender references. As decision-makers, coaches were often blamed for player selection, especially when diversity or "wokeness" was perceived to have influenced team composition. Politicians were more likely to be targeted in connection with religious identity, gender identity, and nationality. Musicians and television broadcasters/journalists were less frequently mentioned overall, but still received hate along lines of gender, when their performance and self-image challenged dominant norms.

Table 15. Proportion of comments (%) containing both identity and individual keywords in the Country Dataset

	Nationality/Country	Gender Identity	Race/Ethnicity	Geography	Religioun
Athlete	32.7	14.4	2.6	0.8	1.3
Referees	1.9	<.001	1.5	.1	<.001
Coach	16.9	5.0	.8	.4	.3
Politician	4.1	2.7	.6	1.3	1.1
Television Broadcaster/Journalist	.2	.2	<.001	.1	<.001
Musician	1.4	3.2	.6	.3	.6
Miscellaneous	3.2	.9	.6	.4	<.001

Table 16 displays the share of identity-individual pairs in comments from the country-based dataset. Results reveal a similar pattern to those in Table 13. Athletes were disproportionately framed through identity lenses, especially nationality and gender, confirming their role as both national symbols and flashpoints for inclusion debates.

Table 16. Proportion of comments (%) containing both identity and individual keywords in the Country Dataset

	Nationality/Country	Gender Identity	Race/Ethnicity	Geography	Religion
Athlete	31.3	13.8	2.4	1.0	.4
Referees	1.8	<.001	1.4	.3	<.001
Coach	10.8	3.1	1.0	.1	.1
Politician	9.7	6.1	0.9	1.8	1.5%
Television Broadcaster/Journalist	.1	.2	<.001	<.001	<.001
Musician	1.2	1.9	.3	.2	.3
Miscellaneous	5.3	1.7	.8	.4	.2

4. Conclusion

This report presents the findings of WP 2 of the RAISE project, which investigated the manifestation of boundary-making and structural racism in user comments on YouTube videos related to four major international sporting events: UEFA Euro 2020, UEFA Euro 2024, UEFA Women's Euro 2022, and the Paris Olympics 2024. Boundary-making is not a simple process, but is constructed at the intersections of race, ethnicity, nationality, religion, gender, and sexuality (Kavanagh et al., 2019; Litchfield et al., 2018; Romero, 2018). Our research efforts, therefore, aimed not only to identify racism, but also hate directed at gender/sexual minorities, nationalities, and religion. Using a combination of NLP techniques, the study analyzed over four million comments across three datasets: event-based, country-based, and institutional channels.

The results suggested that racism persists on social media, despite efforts such as UEFA's call for eliminating racism through campaigns and platforms such as FootbALL (launched in 2023) and Real Scars (launched in 2020), as well as the International Olympic Committee's calls for unity.[2] Athletes emerged as the most visible targets of hate, but within the broader sporting ecosystem, discriminatory behavior in online spaces also targets coaches, politicians, referees, fans, and musicians . Related comments often operated on multiple discursive levels and took the form of insults, homophobic chants, stereotyping, scapegoating, prejudice, and conspiracy theories. These narratives were most prevalent during emotionally intense moments, such as knockout stages and finals, aligning with previous research linking online aggression to high-stakes national competitions. Overall, social media users criticized gameplay, while simultaneously trying to promote their particular views of who belongs and who does not, in both the sport and the nation (Kearns et al., 2023).

The presence of hate speech on official institutional channels, albeit in lower volume, demonstrates that even professionally managed digital spaces are vulnerable to discriminatory narratives. Notably, it was found that none of the sports organizations examined in this report undertook the responsibility of responding to hate comments on their YouTube channels. This absence of engagement further raises concerns about the lack of moderation and the missed opportunity to reinforce inclusive values in sports. Previous research on online anti-racism strategies has identified counter-speech, directly addressing hateful messages through logic, emotional appeals, or humor, as an effective approach. Such responses can serve as a form of social sanctioning, prompting perpetrators to reflect on their behavior and discouraging further hate speech (Coles and Lane, 2023; Baider, 2023). Given these institutions' broad reach and public authority, their silence may be perceived as passive tolerance, resulting in the normalization of discriminatory content. It is therefore essential to encourage proactive moderation and visible institutional engagement to fostering more inclusive digital spaces and uphold the values that these organizations publicly endorse.

---

[2] For UEFA's anti-racism efforts, see https://www.uefa.com/news-media/news/028b-1a765eca983c-8c19960f23bb-1000--tackling-racism-in-european-football/; for the International Olympic Committee's call for unity, see https://www.olympics.com/ioc/news/resolution-of-the-ioc-executive-board-with-regard-to-racism-and-inclusion; (both accessed on 22 April 2025).

References

Abdelrazek, Aly; Eid, Yamna; Gawish, Eman; Medhat, Walaa; and Hassan, Ahmed (2023). “Topic modeling algorithms and applications: A survey.” *Information Systems*, 112, 102131.

Aichner, Thomas; Grünfelder, Matthias, Maurer, Oswin; and Jegeni, Deni (2021). “Twenty-five years of social media: A review of social media applications and definitions from 1994 to 2019.” *Cyberpsychology, Behavior, and Social Networking*, 24(4): 215-222.

Aljarah, Ibrahim; Habib, Maria; Hijazi, Neveen; Faris, Hossam; Qaddoura, Raneem; Hammo, Bassam; Abushariah, Mohammad, and Alfawareh, Mohammad M. (2020). “Intelligent detection of hate speech in Arabic social network: A machine learning approach.” *Journal of Information Science*, 47(4), 483-501.

Alkomah, Fatimah and Ma, Xiaogang (2022) “A literature review of textual hate speech detection methods and datasets.” *Information,* 13.6: 273.

Aloufi, Samah and El Saddik, Abdulmotaleb (2018) “Sentiment identification in football-specific tweets.” *IEEE*, 6: 78609-78621.

Alsagheer, Dana; Mansourifar, Hadi; Dehshibi, Mohammad Mahdi; Shi, Weidong (2022) “Detecting hate speech against athletes in social media.” *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, San Antonio, Texas, USA: 75-81.

Arday, Jason (2023) “The racialised construction of black professional footballers in Engl(ish) football: Love the game, hate the player” In *Critical Issues in Football: A Sociological Analysis of the Beautiful Game*, ed. Will Roberts, Stuart Whigham, Alex Culvin, Daniel Parnell. London: Routledge, 165-173.

Ashraf, Noman; Zubiaga, Arkaitz, and Gelbukh, Alexander (2021) “Abusive language detection in youtube comments leveraging replies as conversational context.” *PeerJ Computer Science,* 7: e742: 1-26.

Avcı, Pelin and Kılınçarslan, Gökmen (2023) The excitement of football kicks off in Qatar: A sample of twitter data analysis.” *The Online Journal of Recreation and Sports,* 12.4: 678-686.

Back, Les and Mills, Kelly (2021) “When you score you’re English, when you miss you’re Black: Euro 2020 and the racial politics of a penalty shoot-out.” *Soundings,* 79.79: 110-121.

Baider, Fabienne (2023) “Accountability issues, online covert hate speech, and the efficacy of counter-speech.” *Politics and Governance,* 11.2: 249-260.

Basile, Valerio; Bosco, Cristina; Fersini, Elisabetta; Nozza, Debora; Patti, Viviana, Pardo, Francisco Manuel Rangel; Rosso, Paolo; Sanguinetti, Manuela (2019) “SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter.” *Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis*, Minnesota, USA: 54–63.

BBC (2021) *Euro 2020 and Covid: How can I watch the final?* https://www.bbc.com/news/uk-57386719. Accessed on 19.03.2025.

Bennett, Samuel (2025) "Soft hate speech and denial of racism at Euro 2020." *Critical Discourse Studies*: 1–15.

Bertoli, Andrew D. (2017) "Nationalism and conflict: Lessons from international sports." *International Studies Quarterly*, 61.4: 835-849.

Bilewicz, Michał and Soral, Wiktor (2020) "Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization." *Political Psychology,* 41(S1): 3-33.

Black, Jack; Fletcher, Thomas; Doidge, Mark; Kearns, Colm; Kilvington, Daniel, Liston, Katie; Lynn, Theo; Rosati, Pierangelo and Sinclair, Gary (2023). "Let the tournament for the woke begin!": Euro 2020 and the reproduction of Cultural Marxist conspiracies in online criticisms of the "take the knee" protest." *Ethnic and Racial Studies*, 47(10): 2036–2059.

Boiy, Erik; Hens, Pieter; Deschacht, Koen; and Moens, Marie-Francine (2007) "Automatic Sentiment Analysis in On-line Text. Ed. Leslie Chan and Bob Martens." *Proceedings of the 11th International Conference on Electronic Publishing*, Vienna, Austria: 349-360.

Brown, Cailin. (2018) "Journalists are gatekeepers for a reason." *Journal of Media Ethics,* 33.2: 94-97.

Burnap, Pete and Williams, Matthew L. (2015) "Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making." *Policy & Internet,* 7.2 (2015): 223-242.

Byun, Unji; Jang, Moonkyoung; and Baek, Hyunmi (2023) "The effect of YouTube comment interaction on video engagement: focusing on interactivity centralization and creators' interactivity." *Online Information Review,*47(6): 1083-1097.

Caldevilla-Domínguez, David; Barrientos-Báez, Almudena; and Padilla-Castillo, Graciela (2023) "Dilemmas between freedom of speech and hate speech: Russophobia on Facebook and Instagram in the Spanish media." *Politics and Governance* 11.2: 147-159.

Camacho-Collados, Jose; Rezaee, Kiamehr; Riahi, Talayeh; Ushio, Asahi; Loureiro, Daniel; Antypas, Dimosthenis; Boisson, Joanne; Espinosa-Anke, Luis; Liu, Fangyu; Eugenio, Martínez-Cámara, Medina, Gonzalo; Buhrmann, Thomas; Neves, Leonardo; and Barbieri, Francesco (2022) "TweetNLP: Cutting-edge natural language processing for social media." *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Abu Dhabi, UAE, 8-49.

Cho, Younghan (2009) "Unfolding sporting nationalism in South Korean media representations of the 1968, 1984 and 2000 Olympics." *Media, Culture & Society,* 31.3 (2009): 347-364.

Cleland, Jamie (2014) "Racism, football fans, and online message boards: How social media has added a new dimension to racist discourse in English football." *Journal of Sport and Social Issues,* 38.5: 415-431.

Coles, Stewart M. and Lane, Daniel S. (2023) "Making the impossible possible? Framing confrontations of racism on social media as norm-setting." *New Media & Society,* 27(4): 2354-2372.

Collier, Kevin (2021) "Twitter's racist algorithm is also ageist, ableist and Islamaphobic, researchers find." *NBC News*, https://www.nbcnews.com/tech/tech-news/twitters-racist-algorithm-also-ageist-ableist-islamaphobic-researchers-rcna1632, accessed on 22.10.2024.

Correa, Teresa and Jeong, Sun Ho (2010) "Race and online content creation: Why minorities are actively participating in the Web." *Information, Communication & Society*, 14(5), 638–659.

Dai, Yaoyao; Gao, Jingjing; and Radford, Benjamin J. (2024) "From fear to hate: Sources of anti-Asian sentiment during COVID-19." *Social Science Quarterly*, *105*(3), 612–635.

Dannenbaum, Claire (2025) "5 facts about Americans and YouTube." *Pew Research Center*, https://www.pewresearch.org/short-reads/2025/02/28/5-facts-about-americans-and-youtube/, accessed on 6 March 2025.

de Leon, Fernanda L.L. and Kim, Sang-Hyun. (2016) "In-group and out-group biases in the marketplace: A field experiment during the World Cup.". https://ssrn.com/abstract=2584414.

Dixon, Kevin; Cashmore, Ellis; and Cleland, Jamie. (2023) "'A little less conversation': An exploration of soccer fan attitudes towards 'the knee'protest and the anti-racism message." *Soccer & Society* 24.5: 698-711.

Dubois, Elizabeth and Reepschlager, Anna (2024) "How harassment and hate speech policies have changed over time: Comparing Facebook, Twitter and Reddit (2005–2020)." *Policy & Internet* 16.3: 523-542.

Duncombe, Constance (2019) "The Polptics of Twitter: Emotions and the power of social media." *International Political Sociology*, 13(4): 409–429.

Ehrmann, Maud; Hamdi, Ahmed; Pontes, Elvys L.; Romanello, Matteo, and Doucet, Antoine (2023). "Named entity recognition and classification in historical documents: A survey." *ACM Computing Surveys*, 56(2), 1-47.

Eime, Rochelle; Harvey, Jack;" Charity, Melanie; and Westerbeek, Hans (2022). "Participation of Australian women and girls in traditionally male-dominated sports 2016-2018." *International Journal of Sport Policy and Politics*, 14(3): 545–561.

Einwiller, Sabine A. and Kim, Sora (2020) "How online content providers moderate user-generated content to prevent harmful online communication: An analysis of policies and their implementation." *Policy & Internet* 12.2: 184-206.

European Parliament (2022) *Flash Eurobarometer 2832: News & Media Survey 2022*.https://search.gesis.org/research_data/ZA7903, accessed on 6 March 2025.

Fenton, Alex; Ahmed, Wasim; Hardey, Mariann; Boardman, Rosy; and Kavanagh, Emma (2024) "Women's football subculture of misogyny: The escalation to online gender-based violence." *European Sport Management Quarterly*, 24(6): 1215-1237.

Forbes, Alison; Edwards, Lisa; and Fleming, Scott (2015) "'Women can't referee': Exploring the experiences of female football officials within UK football culture." *Soccer & Society* 16.4: 521-539.

Frischlich, Lena; Schatto-Eckrodt, Tim; Boberg, Svenja; and Wintterlin, Florian (2021) "Roots of incivility: How personality, media use, and online experiences shape uncivil participation." *Media and Communication*, 9(1): 195-208.

Garcia, Beatriz (2008) "One hundred years of cultural programming within the Olympic Games (1912–2012): Origins, evolution and projections." *International Journal of Cultural Policy*, 14(4): 361–376.

Gelber, Katharine and McNamara, Luke (2016) "Evidencing the harms of hate speech." S*ocial Identities*, 22(3): 324-341.

Glynn, Eleanore and Brown, David H. K. (2022) "Discrimination on football Twitter: The role of humour in the Othering of minorities." *Sport in Society*, 26(8): 1432–1454.

Grootendorst, Maarten (2022) "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794.

Harris, John. (2001) "Playing the man's game: Sites of resistance and incorporation in women's football." World Leisure Journal, 43(4): 22-29.

Hobbs, William; Lajevardi, Nazita; Li, Xinyi; and Lucas, Caleb (2024) "From anti-Muslim to anti-Jewish:Target substitution on fringe social media platforms and the persistence of online and offline hate." *Political Behavior*, *46*(3): 1747-1769.

Hokka, Jenni (2021) "PewDiePie, racism and Youtube's neoliberalist interpretation of freedom of speech." *Convergence* 27.1: 142-160.

Honnibal , Matthew and Montani, Ines (2017) *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*. https://sentometrics-research.com/publication/72/, accessed on 22.04.2025.

Hu , Yan; Chen, Qingyu; Du, Jingcheng; Peng, Xueqing; Keloth, Vipina Kuttichi; Zuo, Xu; Zhou, Yujia; Li, Zehan; Jiang, Xiaoqian; Lu, Zhiyong; Roberts, Kirk; and Xu, Hua (2024) "Improving large language models for clinical named entity recognition via prompt engineering." *Journal of the American Medical Informatics Association*, 31(9): 1812–1820.

Ismer, Sven; Beyer, Manuela; Solms-Baruth, Carolina; and von Scheve, Christian (2015) "International football tournaments, emotional entrainment and the reproduction of symbolic boundaries. A case study in Germany."*Soccer & Society*, 18(4): 554-574.

Jahan, Md Saroar and Oussalah, Mourad (2023) "A systematic review of hate speech automatic detection using natural language processing." *Neurocomputing,* 546: 126232.

Jahns, Martin (2021) "UEFA Euro 2020 and Corona: These are the rules for fans in the 11 host cities." ISPO. https://www.ispo.com/en/know-how/uefa-euro-2020-and-corona-these-rules-apply-fans, accessed on 19 March 2025.

Jehangir, Basra; Radhakrishnan, Saravanan; and Agarwal, Rahul (2023) "A survey on named entity recognition—datasets, tools, and methodologies." *Natural Language Processing Journal,* 3: 100017.

Kaplan, Levi; Gerzon, Nicole; Mislove, Alan; and Sapiezynski, Piotr (2022). "Measurement and analysis of implied identity in ad delivery optimization." In Proceedings of *the 22nd ACM Internet Measurement Conference*: 195-209.

Karayiğit, Habibe; Akdagli, Ali; and Aci, Çiğdem İnan (2022) "Homophobic and hate speech detection using multilingual-bert model on turkish social media." *Information Technology and Control,* 51(2): 356-375.

Kavanagh, Emma; Jones, Ian; & Sheppard-Marks, Lucy (2016) "Towards typologies of virtual maltreatment: sport, digital cultures & dark leisure." *Leisure Studies*, 35(6), 783-796.

Kavanagh, Emma; Litchfield, Chelsea; and Osborne, Jaquelyn (2019) "Sporting women and social media: Sexualization, misogyny, and gender-based violence in online spaces." *International Journal of Sport Communication,*12(4): 552-572.

Kearns, Colm; Sinclair, Gary; Black, Jack; Doidge, Mark; Fletcher, Thomas; Kilvington, Daniel; Liston, Katie; Lynn, Theo; and Rosati, Pierangelo (2023) "A Scoping Review of Research on Online Hate and Sport." *Communication & Sport*, 11(2), 402-430.

Kersting, Norbert (2007) "Sport and national identity: A comparison of the 2006 and 2010 FIFA World Cups." *Politikon*, 34(3), 277-293.

Kilvington, Daniel and Price, John (2019) "Tackling social media abuse? Critically assessing English football's response to online racism." *Communication & Sport,* 7.1: 64-79.

Kilvington, Daniel; Hylton, Kevin; Long, Jonathan; and Bond, Alex (2022) "Investigating online football forums: a critical examination of participants' responses to football related racism and Islamophobia." *Soccer & Society*, 23(8), 849-864.

Koenigstorfer, Joerg; Kolbinger, Otto; and David Gillivray (2023) "Athlete activism, human rights, and the 2020 Olympic Games: a Twitter analysis of Raven Saunders' activism." *Event Management* 27(6): 895-913.

Konikoff, Daniel (2021) "Gatekeepers of toxicity: Reconceptualizing Twitter's abuse and hate speech policies." *Policy & Internet* 13(4): 502-521.

Korostelina, Karina (2014) "Intergroup identity insults: A Social identity theory perspective." *Identity*, 14(3): 214–229.

Kwon, Hye-Jin, Ban; Hyun-Jeong, Jun, Jae-Kyoon; and Kim, Hak-Seon (2021). "Topic modeling and sentiment analysis of online review for airlines." *Information*, 12(2).

Larsson, Anders Olof (2017) "Diversifying likes: Relating reactions to commenting and sharing on newspaper Facebook pages.*" Journalism Practice*, 12(3): 326–343.

Leets, Laura (2022) "Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech." *Journal of Social Issues,* 58.2: 341-361.

Leung, Chun Kai; Ko, Jeremy; Cheung, Anthony; Pun, Boris Lok-Fai; and Ming, Wai Kit (2024) "Borders of belonging: How Chinese nationalism constructs exclusion in the age of populism." *Frontiers in Political Science*, 6: 2.

Levinson, Jordan A.; Greenfield, Patricia M.; and Signorelli, Jenna C. (2020) "A qualitative analysis of adolescent responses to YouTube videos portraying sexual and gender minority experiences: Belonging, community, and information seeking." *Frontiers in Human Dynamics*, 2 (598886).

Li, Jiaxuan and Ning, Yue (2022) "Anti-Asian hate speech detection via data augmented semantic relation inference." *Proceedings of the International AAAI Conference on Web and Social Media*, 16.

Lingiardi, Vittorio; Nicola Carone; Giovanni Semeraro; Cataldo Musto; Marilisa D'Amico; and Silvia Brena (2019) "Mapping Twitter hate speech towards social and sexual minorities: A lexicon-based approach to semantic content analysis." *Behaviour & Information Technology*, 39(7): 711-721.

Lishman, Tom; Shannon, Stephen; McDevitt-Petrovic, Orla; Breslin, Gavin; and Brick, Noel E. (2024) "The impact of abuse experiences and coping responses on distress and mental health outcomes in sports officials." *International Journal of Sport and Exercise Psychology*, 1-22.

Litchfield, Chalsea; Kavanagh, Emma; Osborne, Jaquelyn; and Jones, Ian (2018). "Social media and the politics of gender, race and identity: the case of Serena Williams." *European Journal for Sport and Society,* 15(2): 154–170.

MacInnes, Paul and Duncan, Pamela (2021) "'Unforgivable': Gareth Southgate on racist abuse of England players." *The Guardian*. https://www.theguardian.com/football/2021/jul/12/gareth-southgate-racist-abuse-of-england-players-unforgiveable, accessed on 13 March 2025.

Madden, Amy; Ruthven, Ian; and McMenemy, David (2013) "A classification scheme for content analyses of YouTube video comments." *Journal of Documentation*, 69(5): 693-714.

Marivoet, Salomé (2006) "UEFA Euro 2004 Portugal: The social construction of a sports mega-event and spectacle." *The Sociological Review,* 54(2): 127-143.

Matamoros Fernandez, Ariadna (2018) "Inciting anger through Facebook reactions in Belgium: The use of emoji and related vernacular expressions in racist discourse." *First Monday,* 23.9: 1-20.

Matamoros-Fernández, Ariadna (2017) "Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube." *Information, Communication & Society*, 20(6): 930–946.

Matamoros-Fernández, Ariadna; Gray, Joanne E.; Bartolo, Louisa; Burgess, Jean; and Suzor, Nicolas (2021) "What's up next? Investigating algorithmic recommendations on youtube across issues and over time." *Media and Communication*, 9(4): 234-249.

Mauro, Max (2024) "Contrasting media representations of race and national identity: The case of England and Italy at the Union of European Football Associations Euro 2020." *European Journal of Cultural Studies*, 28(2): 472-492.

McVeigh, Niall and Hall, Rachel (2021) “Boris Johnson condemns ‘appalling’ racist abuse of England players.” *The Guardian* https://www.theguardian.com/football/2021/jul/12/fa-condemns-racist-abuse-england-players-social-media-euro-2020-final, accessed on 13 March 2025.

Meriläinen, Jari-Mikko (2024) “The role of gender in hate speech targeting politicians: Evidence from Finnish Twitter.” *International Journal of Politics, Culture, and Society*: 1-27.

Miranda, Sandra; Gouveia, Célia; Di Fátima, Branco; and Antunes, Ana Cristina (2023) “Hate speech on social media: Behaviour of Portuguese football fans on Facebook.” *Soccer & Society*, 25(1): 76–91.

Mujahid, Muhammad; Lee, Ernesto; Rustam, Furqan; Washington, Patrick B.; Ullah, Saleem; Reshi, AijazA.; and Ashraf, Imran (2021) “Sentiment analysis and topic modeling on tweets about online education during COVID-19.” *Applied Sciences*, 11(18): 8438.

Mullah, Nanlir Sallau; Zainon, Wan Mohd Nazmee Wan; and Wahab, Mohd Nadhir Ab (2024) “Transfer learning approach for identifying negative sentiment in tweets directed to football players.” *Engineering Applications of Artificial Intelligence*, 133: 1-11.

Myers West, Sarah (2018) “Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms.” *New Media & Society*, 20(11): 4366-4383.

Nadkarni, Prakash M.; Ohno-Machado, Lucila; and Chapman, Wendy W. (2011) “Natural language processing: An introduction.” *Journal of the American Medical Informatics Association* 18(5): 544-551.

Nägel, Christof; Kros, Mathijs; and Davenport, Ryan (2024) “Three lions or three scapegoats: Racial hate crime in the wake of the Euro 2020 final in London.” *Sociological Science,* 11: 579-599.

Näsi, Matti; Räsänen, Pekka; Hawdon, James; Holkeri, Emma; Oksanen, Atte (2015) “Exposure to online hate material and social trust among Finnish youth” *Information Technology & People*, 28(3): 607-622.

O’Neill, Francis; Dickson, Geoff; Ströbel, Tim; and Thompson, Ashleigh-Jane (2025) “Elite athlete activism, advocacy, and protest: A scoping review.” *European Sport Management Quarterly*, 25(1): 145-173.

Oshiro, Kristi F.; Weems, Anthony J.; and Singer, John N. (2020) “Cyber racism toward black athletes: A critical race analysis of texags.com online brand.” community. communication & sport, 9(6): 911-933.

Pookpanich, Peerat and Siriborvornratanakul, Thitirat (2024) “Offensive language and hate speech detection using deep learning in football news live streaming chat on YouTube in Thailand.” *Social Network Analysis and Mining,* 14(1): 18.

Poulton, Emma, and Durell, Oliver (2016) “Uses and meanings of ‘Yid’in English football fandom: A case study of Tottenham Hotspur Football Club.” *International Review for the Sociology of Sport,* 51.6: 715-734.

Reynolds, CJ and Hallinan, Blake (2024) “User-generated accountability: Public participation in algorithmic governance on YouTube.” New Media & Society, 26(9): 5107-5129.

RAISE

Rolf, Jan Niklas (2023) "Olympism, cosmopolitanism, nationalism: The many face (t) s of the Olympics."*Journal of Political Ideologies,* 28(1): 102-120.

Romero, Mary. *Introducing Intersectionality*. New Jersey: John Wiley & Sons, 2024.

Rosenzweig, Leah R., and Zhou, Yang-Yang (2021) "Team and nation: Sports, nationalism, and attitudes toward refugees." *Comparative Political Studies* 54(12): 2123-2154.

Rotman, Dana and Preece, Jennifer (2010) "The'WeTube'in YouTube: Creating an online community through video sharing." *International Journal of Web Based Communities*, 6(3): 317-333.

Rudwick, Stephanie and Schmiedl, Martin (2023) "'It's not our problem': Czech online discourse on kneeling in football and Black Lives Matter." *Sport in Society* 26(10): 1701-1722.

Sanderson, Jimmy (2013) "From loving the hero to despising the villain: Sports fans, Facebook, and social identity threats." *Mass Communication and Society*, 16(4), 487-509.

Sanderson, Jimmy; Frederick, Evan; Stocz, Mike (2016) "When athlete activism clashes with group values: Social identity threat management via social media." *Mass Communication and Society*, 19(3): 301-322.

Scissors, Lauren; Burke, Moira; and Wengrovitz, Steven (2016) "What's in a Like? Attitudes and behaviors around receiving Likes on Facebook." *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*.

Seijbel, Jasmin; van Sterkenburg; Jacco; and Oonk, Gijsbert (2022) "Expressing rivalry online: Antisemitic rhetoric among Dutch football supporters on Twitter." *Soccer & Society*, 23(8); 834-848.

Seijbel, Jasmin; van Sterkenburg; Jacco; and Spaaij, Ramón (2023) "Online football-related antisemitism in the context of the COVID-19 pandemic: A multi-method analysis of the Dutch Twittersphere." *American Behavioral Scientist,* 67(11): 1304-1321.

Siapera, Eugenia and Viejo-Otero, Paloma (2021) "Governing hate: Facebook and digital racism." *Television & New Media*, 22(2): 112-130.

Soral, Wiktor; Bilewicz, Michał; Winiewski, Mikołaj (2018) "Exposure to hate speech increases prejudice through desensitization." *Aggressive Behavior*, 44(2): 136-146.

Soral, Wiktor; Liu, James; and Bilewicz, Michał (2020) "Media of contempt: Social media consumption predicts normative acceptance of anti-Muslim hate speech and islamoprejudice." *International Journal of Conflict and Violence*, 14: 1-13.

Ștefăniță, Oana and Buf, Diana-Maria (2021) "Hate speech in social media and its effects on the LGBT community: A review of the current research." *Romanian Journal of Communication and Public Relations* 23.1: 47-55.

Stirling, Ashley E. (2013) "Understanding the use of emotionally abusive coaching practices." *International Journal of Sports Science & Coaching,* 8(4): 625-639.

Thelwall, Mike (2018) "Social media analytics for YouTube comments: Potential and limitations." *International Journal of Social Research Methodology*, 21(3): 303-316.

Turki, Turki and Sanjiban, Sekhar Roy (2022) "Novel hate speech detection using word cloud visualization and ensemble learning coupled with count vectorizer." *Applied Sciences* 12.13: 6611.

Vincent, John; Kian, Edward M.; Pedersen, Paul M.; Kuntz, Aaron; and Hill, John S. (2010) "England expects: English newspapers' narratives about the English football team in the 2006 World Cup." *International Review for the Sociology of Sport*, 45(2): 199-223.

Voorveld, Hilde A. M.; van Noort, Guda; Muntinga, Daniël G.; and Bronner, Fred (2018) "Engagement with social media and social media advertising: The differentiating role of platform type." *Journal of Advertising*, 47(1), 38-54.

Vujičić Stanković, Staša and Mladenović, Miljana (2023) "An approach to automatic classification of hate speech in sports domain on social media." *Journal of Big Data* 10(109): 1-16.

Wang, Haifeng; Wu, Hua; He, Zhongjun; Huang, Liang; and Church, Kenneth Ward (2022) "Progress in machine translation." *Engineering*, 18: 143-153.

Wann, Daniel L.; Culver, Zachary; Akanda, Rubaba; Daglar, Melek; De Divitiis, Carla; and Smith, Anthony (2005). "The effects of team ıdentification and game outcome on willingness to consider anonymous acts of hostile aggression." *Journal of Sport Behavior*, 28(3): 282-294.

Wilson, Richard Ashby (2022) "Digital authoritarianism and the global assault on human rights." *Human Rights Quarterly*, 44.4: 704-739.

Wimmer, Andreas (2008). "Elementary strategies of ethnic boundary making." *Ethnic and Racial Studies*, *31*(6), 1025-1055.

Xiang, Ning; Wang, Limao; Zhong, Shuai; Zheng, Chen; Wang, Bo; and Qu, Qiushi (2021). "How does the world view China's carbon policy? A sentiment analysis on twitter data." *Energies* 14(22): 7782.

YouTube (2025) *Growing Your Channel*. https://www.youtube.com/creators/how-things-work/content-creation-strategy/, accessed on 6 March 2025.

Zampieri, Marcos; Malmasi, Shervin; Nakov, Preslav; Rosenthal, Sara; Farra, Noura; and Kumar, Ritesh (2019) "Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval)." *arXiv preprint arXiv:1903.08983*.

RAISE

Appendix 1. Distribution of Languages in the Event Dataset

Keywords											
		euro 2020		euro 2022		euro 2024		Paris Olympics 2024		Paris2024	
	Language	N	%	N	%	N	%	N	%	N	%
1	af	2,345	.840	150	.871	2,106	1.146	345	.934	1,370	.888
2	ar	182	.065	4	.023	462	0.251	37	.1	126	.082
3	bg	90	.032	12	.07	95	0.052	19	.051	246	.16
4	bn	12	.004	0	0	10	0.005	25	.068	5	.003
5	ca	1,828	.655	66	.383	1,573	0.856	137	.371	1,334	.865
6	cs	351	.126	17	.099	243	0.132	50	.135	182	.118
7	cy	4,080	1.461	162	.941	2,307	1.256	360	.974	1,569	1.017
8	da	1,950	.698	81	.470	1,669	0.908	203	.549	674	.437
9	de	5,231	1.873	648	3.763	5,021	2.733	519	1.404	2,811	1.823
10	el	534	.191	0	0	38	0.021	4	.011	173	.112
11	en	144,757	51.840	12,314	71.51	85,561	46.574	23,378	63.257	78,143	5.67
12	es	6,963	2.494	334	1.94	7,102	3.866	175	.474	6,460	4.189
13	et	1,471	.527	57	.331	1,860	1.012	486	1.315	1,034	.670
14	fa	24	.009	1	.006	312	0.170	11	.03	64	.041
15	fi	1,477	.529	46	.267	1,952	1.063	174	.471	712	.462
16	fr	6,359	2.277	773	4.489	6,466	3.520	414	1.12	16,037	1.399
17	gu	0	0	0	0	0	0	4	.011	0	0
18	he	11	.004	9	.052	7	0.004	1	.003	15	.01

19	**hi**	11	.004	0	0	1	0.001	967	2.617	327	.212
20	**hr**	1,158	.415	156	.906	1,078	0.587	212	.574	618	.401
21	**hu**	1,238	.443	46	.267	610	0.332	69	.187	546	.354
22	**id**	7,093	2.54	95	.552	5,262	2.864	2,048	5.542	3,067	1.989
23	**it**	24,149	8.648	375	2.178	3,452	1.879	256	.693	2,605	1.689
24	**ja**	79	.028	8	.046	59	0.032	47	.127	433	.281
25	**kn**	0	0	0	0	2	0.001	1	.003	1	.001
26	**ko**	664	.238	6	.035	56	0.030	56	.152	1,018	.66
27	**lt**	488	.175	16	.093	670	0.365	51	.138	299	.194
28	**lv**	173	.062	5	.029	273	0.149	22	.06	126	.082
29	**mk**	46	.016	11	.064	38	0.021	10	.027	135	.088
30	**ml**	30	.011	0	0	5	0.003	0	0	5	.003
31	**mr**	1	0	0	0	0	0	75	.203	22	.014
32	**ne**	1	0	0	0	3	0.002	48	.13	14	.009
33	**nl**	2,808	1.006	126	.732	3,336	1.816	301	.814	1,227	.796
34	**no**	1,852	.663	70	.407	1,805	0.983	177	.479	852	.552
35	**pa**	0	0	0	0	0	0	7	.019	13	.008
36	**pl**	11,703	4.191	48	.279	1,929	1.050	92	.249	1,525	.989
37	**pt**	6,289	2.252	111	.645	3,439	1.872	157	.425	3,819	2.476
38	**ro**	3,411	1.222	56	.325	2,004	1.091	228	.617	1,394	.904
39	**ru**	478	.171	45	.261	442	0.241	274	.741	1,642	1.065
40	**sk**	563	.202	18	.105	296	0.161	67	.181	452	.293
41	**sl**	1,108	.397	61	.354	1,002	0.545	191	.517	570	.370

42	**so**	4,518	1.618	151	.877	3,234	1.760	1,187	3.212	2,656	1.722
43	**sq**	470	.168	26	.151	477	0.260	229	.62	236	.153
44	**sv**	844	.302	33	.192	740	0.403	107	.29	373	.242
45	**sw**	1,265	.453	36	.209	1,619	0.881	644	1.743	1,562	1.013
46	**ta**	1	0	0	0	1	0.001	10	.027	3	.002
47	**te**	0	0	0	0	1	0.001	7	.019	1	.001
48	**th**	42	.015	4	.023	31	0.017	1	.003	57	.037
49	**tl**	3,310	1.185	191	1.109	2,910	1.584	540	1.461	1,986	1.288
50	**tr**	15,257	5.464	317	1.841	16,457	8.958	246	.666	5,921	3.839
51	**uk**	72	.026	1	.006	49	0.027	14	.038	168	.109
52	**unknown**	10,199	3.652	505	2.933	13,682	7.448	2,008	5.433	8,022	5.202
53	**ur**	14	.005	1	.006	45	0.024	102	.276	6	.004
54	**vi**	2,209	.791	28	.163	1,909	1.039	113	.306	513	.333
55	**zh-cn**	27	.01	0	0	9	0.005	49	.133	672	.436
56	**zh-tw**	2	.001	0	0	0	0	2	.005	378	.245
	Total Comments	279,238	100	17,220	100	183,710	100	36,957.00	100	154,219	100