

Sportswashing Narratives on YouTube: An NLP Study of Sentiment & Misinformation around Gulf Investments.

Student Name	Supervisor Name	Programme of Study
Adil Khan	Michael Schlichtkrull	Big Data Science

ABSTRACT

This study investigates public responses to Gulf state investments in global sport—commonly termed sportswashing—by analysing 72,000 YouTube comments across events like the FIFA World Cup, LIV Golf, and club takeovers. This project used NLP, large language models, and fact-checking tools to analyse sentiment, misinformation, and alignment with video narratives. Findings revealed dominant negativity, especially around Saudi Arabia and human rights, with misinformation clustering in emotionally charged, dissenting comments. A fine-tuned sentiment model showed strong predictive performance, indicating consistent discourse patterns across events. The research offers a scalable framework for examining public opinion on politicized sport and highlights both the potential and challenges of applying LLMs to real-world online discourse.

I. INTRODUCTION

In recent years, the growing presence of Gulf nations in global sport has raised important questions about reputation, politics, and power. Countries like Saudi Arabia, Qatar, and the UAE have made high-profile investments in football, Formula 1, and golf, often framed as strategic efforts to build global influence and reshape international perception. This phenomenon, commonly referred to as sportswashing, sits at the intersection of sport, soft power, and image management.

While these investments are frequently discussed in the media and policy circles, it's just as important to understand how everyday people respond to them. Social platforms like YouTube host thousands of comments reacting to promotional videos, match highlights, press conferences, and documentaries tied to these events. These reactions often reveal how the public feels, what concerns they raise, and whether they believe the narratives being presented.

This thesis investigates public sentiment and misinformation around Gulf sportswashing through YouTube comments. Using NLP, large language models (LLMs), and fact-checking tools, I collected 110,000 comments from major events in Saudi Arabia, Qatar, and the UAE. After filtering, 72,000 comments were analysed across stages including comment rewriting, claim extraction, article retrieval, and LLM-based verification. The goal is to classify sentiment, detect misinformation, and understand how people engage with official narratives across different countries and sports.

By combining quantitative analysis (such as sentiment scores and claim verification outcomes) with qualitative insight (like recurring themes and agreement with transcripts), this research offers a deeper look at how public discourse reflects and often resists efforts to control image through sport.

II. RELATED WORK

A. YouTube Comments as a Data Source

YouTube has become a widely used platform for studying online discourse, with comment sections offering a large, unstructured source of public opinion. Siersdorfer et al. (2010) analysed over 6 million comments to understand sentiment trends and how they relate to video categories and user engagement. More recent studies have focused on misinformation, such as Suter et al. (2022), who collected 3.5 million comments related to COVID-19 and classified them to detect verifiably false claims. Researchers have also used YouTube comments to examine topics like hate speech, tourism perception, and political sentiment. These studies often highlight the richness of the data, especially compared to more character limited platforms like X.

B. Sportswashing & Gulf Investment in Sport

The term “Sportswashing” has been used to describe how states use sport to improve their image and distract from human rights issues or authoritarian practices. Ganji (2023) explored this in the context of Qatar, Saudi Arabia, and the UAE, describing how sports are used as part of broader reputation management strategies. Grix et al. (2023) examined the political function of sportswashing and its reliance on global sports institutions willing to engage with authoritarian regimes. Brannagan and Giulianotti (2015) introduced the concept of “soft disempowerment” to describe cases where sportswashing efforts can backfire due to increased global scrutiny. The 2022 FIFA World Cup in Qatar has been a major focus in this area, with Gerschewski et al. (2024) running large-scale public opinion studies that showed how framing and media context influenced international perceptions of Qatar during the event.

C. NLP Techniques for Social Media Analysis

Natural Language Processing has been widely applied to social media data for tasks like sentiment analysis, topic modelling or stance detection. Early studies often used lexicon-based approaches, such as the NRC Emotion Lexicon, or unsupervised models like Latent Dirichlet Allocation to identify dominant themes. With the rise of deep learning, models like Bert and RoBERTa have become popular for more context-sensitive analysis. These transformer-based models are capable of handling informal language and code-switching often found in online discourse, making them suitable for tasks involving public opinion, emotions, and online debates. NLP has also been applied in multilingual contexts and cross-platform studies, especially in political communication and crisis analysis.

D. Misinformation Detection & Automated Fact-Checking

There is a growing body of research dedicated to automated fact-checking and misinformation detection. Standard approaches typically involve detecting check-worthy claims, retrieving supporting or refuting evidence, and generating a verification verdict. Schlichtkrull et al. (2021) provides a detailed survey of fact checking systems, outlining common methods used for claim detection, evidence retrieval, and stance classification. More recent work by Schlichtkrull et al. (2023) critiques the design of these systems, highlighting the importance of clarifying their intended use cases and epistemic role they are meant to serve. Alongside this, newer datasets like FEVEROUS have pushed the field towards more realistic verification tasks involving both structured and unstructured evidence. These developments reflect an ongoing shift from isolated content classification towards more context-aware systems, where trust, explanation, and practical usability are becoming central design concerns.

E. Large Language Models & Social Data Analysis

Large Language Models (LLMs), particularly transformer-based systems like GPT-3 and GPT-4, have demonstrated strong performance across NLP tasks including sentiment classification, text generation, and zero-shot inference. Researchers have explored their use in misinformation detection by prompting models to explain, verify, or refute claims using natural language outputs. While these models have achieved impressive results in benchmarks, studies have also raised concerns about their reliability. For instance, Islam et al. (2024) highlight that hallucinations—plausible-sounding but ungrounded or incorrect outputs—pose significant challenges to LLM deployment in real-world applications. These hallucinations can undermine the trustworthiness of AI systems, especially in sensitive domains like healthcare and finance. To address this, researchers are investigating mitigation strategies such as Retrieval-Augmented Generation (RAG) and reinforcement learning with human feedback (RLHF) to improve the factual accuracy and reliability of LLM outputs. Despite these efforts, human oversight remains critical to ensure responsible use in misinformation detection and other high-stakes contexts.

III. METHODOLOGY

A. Data Collection

Data for this project was collected directly from YouTube using the official YouTube Data API. YouTube was chosen as the primary data source after considering other platforms like X (formerly Twitter) and Reddit. While X has been widely used in social media research, changes to its API pricing made large-scale data collection costly and restrictive. Reddit, though useful for structured discussions, was less effective for directly targeting event-specific content related to gulf investments in sport. YouTube on the other hand, allowed me to search directly for relevant videos and collect comment threads tied to specific sporting events, promotional content, and press coverage. This made it ideal for capturing both the official narrative and the spontaneous public response in one place, making it not only the most

accessible option but also the most aligned with the aims of this project.

Videos were carefully selected using a combination of event-specific search terms and manual verification to ensure topical relevance. For each video, the API collected essential metadata, including:

Video ID	Comment ID	Reply Count
Video Title	Comment	Video Category
Channel Name	Author	
Date	Likes	

The dataset covers several high-profile cases, including:

- **Football Ownership:** Manchester City (UAE), Paris Saint-Germain (Qatar), Newcastle United (Saudia Arabia)
- **Major Tournaments:** FIFA World Cup 2022 (Qatar), Formula 1, Saudi Pro League growth
- **Gulf-led Initiatives:** LIV Golf (Saudia Arabia) and broader multi-sport hosting efforts in the region

The collected data was saved by event and later merged into a consolidated dataset containing around 110,000 YouTube comments. In addition to the comments, I also retrieved full video transcripts from the same set of videos using the YouTube transcript API. These transcripts provided a word-for-word record of the spoken content in each video, enabling deeper comparison between the original messaging and the audience’s response. Each transcript was stored alongside its corresponding video metadata, making it possible to later evaluate whether user comments were influenced by the video content and whether they aligned or disagreed with key points from the transcript.

B. Preprocessing & Filtering

Before any analysis could be carried out, the collected comments and transcripts had to be cleaned, standardized, and enriched with a range of labels to support downstream tasks. This stage combined manual scripting with LLM-powered classification to create a structured, reliable dataset.

I started by rewriting each comment using GPT-4o Mini via one-by-one API calls to improve clarity and reduce noise from unstructured or typo-ridden comments. This made the data cleaner and reduced misclassification risk. However, the process was extremely slow, taking several days to complete. After reviewing OpenAI’s documentation, I transitioned to **batch processing**, which was far more efficient and better suited for large-scale workflows. I restructured the pipeline to submit large batches of comments – sometimes processing thousands at a time, including batches of 1,000, 2,000, and even 5,000 depending on the task and model constraints. This allowed me to classify tens of thousands of comments in parallel while keeping costs and processing time manageable. The first batch processing task was to filter out **non-thesis-relevant** comments. I used GPT-4o Mini to determine whether each rewritten comment was related to themes like sportswashing, Gulf states, human rights, or corruption. Comments were labelled as **Yes**, **No**, or **Maybe**, depending on how confidently the model could classify them. For Maybe comments, I used sentence embeddings (SBERT) to extract the top 5 to 8 most relevant lines from the transcript. This let me compare each comment against a condensed version of the transcript, which saved on token usage while

still providing enough context to decide. I later applied this same technique for both the Influenced by Transcript and Agree with Transcript columns, again using GPT-4o for the final classification after narrowing down the transcript using SBERT.

This same pipeline was used for all other classification tasks. Simpler ones like sentiment or emotion were handled by GPT-4o Mini, while more complex tasks like claim detection or agreement used GPT-4o. **As shown in the figure at the end of the Methodology**, each step involved submitting the rewritten comment with a task-specific prompt. For claim detection, the model was asked: “Does this comment contain a verifiable factual claim? Reply ONLY with 1 = yes, 0 = no.” All prompts are listed in Appendix A. This process produced a labelled dataset of over 72,000 cleaned comments.

Column Name	Purpose/ Description	Model Used
Rewritten comment	Cleaned and standardized version of the original comment for consistency	GPT-4o Mini
Thesis Relevant	Determines whether the comment is related to Gulf sportswashing topics	GPT-4o Mini
Comment Category	Assigns topic tags like sportswashing, human rights, corruption, etc.	GPT-4o Mini
Country	Identifies the Gulf state (e.g. Saudi Arabia, Qatar, UAE) associated with the comment	GPT-4o Mini
Event Type	Classifies the kind of event (e.g. ownership, tournament, initiative) linked to the comment	GPT-4o Mini
Sentiment	Classifies the tone of the comment as positive, neutral, or negative	GPT-4o Mini
Emotion Class	Labels the dominant emotion (e.g. anger, worry, optimism) in the comment	GPT-4o Mini
Fact or Opinion	Identifies whether the comment expresses a factual claim or an opinion	GPT-4o Mini
Claim Detection	Flags whether the comment contains a check-worthy factual claim	GPT-4o
Influenced by Transcript	Detects whether the comment responds directly to the content of the video	GPT-4o
Agreed with Transcript	Identifies whether the commenter agrees, disagrees, or stays neutral	GPT-4o

Table 1: Pre-processed columns

C. Sentiment & Emotion Classification

Once sentiment and emotion labels were assigned to each comment, I used them to explore how public responses varied across different countries, events, and phases. Sentiment was grouped by country (e.g. Saudi Arabia, Qatar, UAE) and by event type (such as the FIFA World Cup or LIV Golf) to see how overall tone shifted depending on the context. I also tracked how sentiment changed over time — comparing pre-event, event window, and post-event periods — to identify spikes in negativity or optimism tied to specific announcements or matches. In parallel, the emotion labels provided a more detailed view of how people felt. For

example, I compared how often emotions like anger or worry appeared in response to events linked to human rights concerns or corruption narratives. I also used emotion patterns to help explain certain claim trends or engagement spikes — especially when negative emotions overlapped with misinformation or politically sensitive themes. Together, sentiment and emotion scores helped map out the public mood and how it evolved across different parts of the dataset.

D. Claim Detection and Article Verification

To identify comments that made factual claims worth checking, I used the output from the Claim Detection column to isolate relevant rows for verification. This column flagged whether a comment contained a check-worthy factual claim, and only those marked as claims were passed into the next stage. I initially tested the Google Fact Check API to support verification, but it turned out to be too limited for my dataset. After running it on a sample of over 10,000 claims, it returned only one matched result, making it unfit for large-scale use. I also considered using benchmark datasets like FEVER (Thorne et al., 2018), but since many of the claims in my dataset referred to real-world events happening in the last few years, I decided a real-time search approach would be more appropriate. I eventually used GPT-4o deployed via **Azure Foundry**, combined with **Bing Search grounding**, which allowed the model to pull in live web evidence during verification.

Claims were submitted one by one through the Azure API. While this method was slower, it gave me more control over the verification process and helped maintain accuracy and traceability in the outputs. Each API call returned a verdict, either Likely True, Likely False, Unverifiable, or Opinion or Speculation, along with a short explanation and a list of supporting sources. The claim categories used in this stage were drawn from the thematic tags assigned earlier using GPT-4o Mini. I focused this process on claims grouped by their assigned category, such as corruption, human rights, sportswashing, or financial ethics, to ensure a balanced spread of topics during verification. These claim verdicts were later used to explore misinformation patterns across the dataset, compare claim outcomes by event and theme, and examine how factual accuracy aligned with sentiment and engagement trends.

E. Transcript Based Analysis

Once the Influenced by Transcript and Agreed with Transcript labels were assigned, I used them to examine how viewers interacted with the video content itself. I compared rates of transcript influence across different countries, event types, and topics to see which kinds of videos generated more direct engagement. I also analysed how often comments agreed or disagreed with the messaging in each video, and whether those patterns varied depending on the event category or narrative theme.

I also used transcript-based labels alongside sentiment, emotion, and claim data to examine how alignment with the video affected tone and factual accuracy. For instance, I tested whether disagreeing comments were more prone to misinformation, or if transcript influence was stronger in

neutral vs polarized sentiment. This clarified how user reactions tracked with or resisted the video’s narrative.

F. Temporal Tagging

Each comment in the dataset was linked to a specific time window — either pre-event, event window, or post-event — based on the upload date of the video it was posted under. These time periods were defined manually for each event using the associated metadata, allowing for consistent segmentation across different sports and countries.

I used these temporal tags to group sentiment, emotion, and misinformation labels over time. This helped identify whether public opinion shifted as events unfolded, and whether certain themes or types of misinformation appeared more frequently during or after high-profile moments. It also made it possible to compare whether engagement levels spiked during the event itself or were more reactive in the lead-up or aftermath. By combining time segmentation with the other classification outputs, I was able to track how reactions evolved across different phases of each event and spot patterns that wouldn’t have been visible in a static snapshot of the data.

G. Network Analysis

To explore how viewer responses clustered around different types of content, I built a bipartite graph connecting videos and the comments associated with them. Each video acted as a central node, and each comment was linked to the video it was posted under using the Video ID column. This structure allowed me to treat videos as hubs of discourse and observe how reactions varied by country, sport, or event type. It also helped identify where key narratives — like sportswashing, corruption, or human rights — were most concentrated.

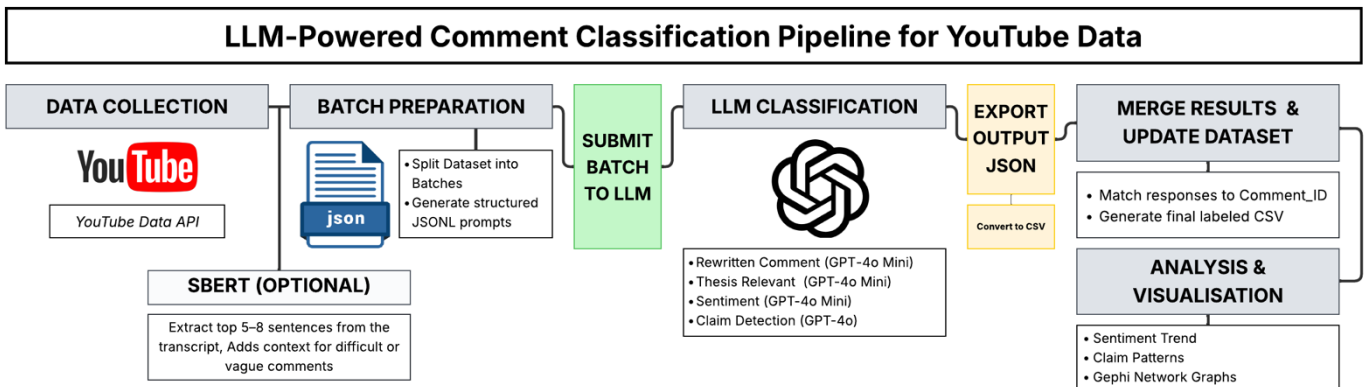
To dig deeper, I used node-level metadata to measure and visualize differences in how videos were received. I calculated a basic engagement score for each comment ($\text{Likes} + 2 \times \text{Replies Count}$) to highlight those with greater visibility.

This approach aligns with industry practice, such as Brandwatch (2022), which recommends weighting comments more heavily than likes to reflect deeper engagement. I also coloured or sized comment nodes in Gephi based on sentiment, emotion, claim verdict, or category to explore how narratives or misinformation clustered around certain videos. On the video side, I looked at which videos showed higher densities of negative sentiment, disagreement with the transcript, or Likely False claims. While not a traditional interaction network, this still revealed how discourse formed around specific content, and which events triggered the most divided or emotionally intense responses.

H. Sentiment Prediction (2034 World Cup)

As a final exploratory step, I attempted to estimate how public sentiment might develop around future Gulf-hosted sporting events, using the 2034 FIFA World Cup (expected to be hosted by Saudi Arabia) as a case study. While this wasn’t a core focus of the project, it served to reflect on whether patterns observed in the current dataset could be used to anticipate public discourse going forward.

In addition to Saudi-led events — including LIV Golf, the Saudi Pro League, and the Newcastle United takeover — I incorporated findings from the 2022 FIFA World Cup in Qatar, which offered a comparable regional and narrative context. To support this analysis, I fine-tuned a DistilBERT-based sentiment classification model on labelled YouTube comments, with sentiment mapped to three classes. Rather than applying the model to future data, I used its outputs alongside emotion trends, misinformation clustering, and transcript disagreement to explore how similar reactions might emerge around Saudi Arabia 2034. The aim was not to produce a forecast, but to assess whether sentiment across such events follows identifiable and potentially predictable patterns.



IV. RESULTS & DISCUSSION

This section presents the main findings from the analysis. Each result links back to the key components of the methodology — including sentiment trends, misinformation patterns, transcript engagement, and thematic distribution. I also discuss how these findings varied across different countries, events, and time windows, and what they suggest about public reactions to Gulf state involvement in global sport. Visualizations and summary tables are used where relevant to highlight the patterns that emerged.

A. Sentiment & Emotion Trend

To begin the analysis, I examined the overall distribution of sentiment and emotion across all 72,000 comments. The dataset showed a clear lean toward negative reactions, with positive sentiment forming a smaller proportion of the total. Table 2 provides a summary of the sentiment categories alongside the six most frequently occurring emotion classes.

Sentiment	Count	Emotion Class	Count
Positive	8,765	Disapproval	32,610
Neutral	21,092	Anger	21,761
Negative	41,583	Joy	6,481
		Neutral	5,562
		Sadness	3,276
		Hope	1,643

Table 2: Sentiment & Emotion Distribution in YouTube comments

Emotions like disapproval and anger appeared more often than negative sentiment alone, since emotion captures dominant feelings even in neutral-toned comments. Users may express frustration while maintaining a balanced stance. Meanwhile, less frequent emotions like joy and hope show that reactions weren't entirely negative, offering a more layered picture than sentiment alone.

Country	Negative	Neutral	Positive
Qatar	29,315	14,163	6,476
Saudi Arabia	9,608	5,541	1,985
UAE	2,589	1,361	294

Table 3: Sentiment Breakdown by Country

Sentiment varied noticeably across countries. Qatar received the highest volume of comments, with a strong lean toward negativity but also a significant share of positive sentiment, particularly around the World Cup. Emotionally, Qatar-related comments showed high levels of anger and disapproval, but also signs of hope and joy, suggesting that while criticisms were common, many viewers still appreciated the event's cultural or sporting aspects. Saudi Arabia, on the other hand, drew consistently negative reactions, with far fewer positive or neutral responses. The tone here was dominated by frustration, especially in relation to human rights and sportswashing concerns. UAE-linked content, mostly referring to Manchester City ownership, also showed high levels of disapproval, though the overall volume was lower. Across all three, public reaction was shaped not just by the sporting context but also by each country's broader reputation and political associations.

Country	Negative	Neutral	Positive
FIFA World Cup 2022	29,280	14,017	6,467
Saudi Pro League	2,901	2,070	794
Manchester City Ownership	3,133	1,348	276
Newcastle United Ownership	2,681	1,458	487
LIV Golf	2,064	1,053	314
Gulf Multi-Sport hosting	456	267	185
Formula 1	335	193	59
PSG Ownership	237	185	38

Table 4: Sentiment Distribution by Event

Reactions also differed depending on the event. The FIFA World Cup 2022 saw the highest engagement, with a mix of strong criticism and praise. While most comments were negative, a sizable number expressed joy and hope, often tied to regional pride or the scale of the tournament. This emotional range showed that even controversial events can inspire positive responses. Events like LIV Golf and the Saudi Pro League drew more one-sided reactions — dominated by anger, disapproval, and scepticism toward Gulf investment in sport. Ownership cases like Manchester City and Newcastle United also attracted criticism, though some responses were more mixed, with a few neutral or even amused tones. Less-discussed events like Formula 1 and PSG ownership followed a similar trend, though with lower emotional intensity. Overall, the emotional weight appeared linked not only to the event itself, but to its visibility or political sensitivity, a pattern supported by the disproportionately high levels of negative sentiment in events like the FIFA World Cup and LIV Golf (see Table 4).

B. Claim Detection & Verification

As this process relied on a language model, some misclassifications are likely, particularly in comments using sarcasm, indirect phrasing, or informal language (see Appendix B for examples). As shown in Table 5, most comments were classified as opinion-based, with only a small proportion identified as fully factual. However, when applying claim detection, a much larger subset of 26,537 comments was flagged as containing check-worthy claims. This suggests that users frequently embedded factual assertions within statements otherwise subjective in tone.

Label	Count	Label	Count
Opinion	67,678	Claim	26,537
Factual	3,763	No Claim	44,904

Table 5: Distribution of Factuality and Claim Detection Labels

The most common verdict was Opinion or Speculation, followed by Likely True, with fewer marked as Unverifiable or Likely False. This reflects the informal, often rhetorical nature of online comments, but also shows many users engaged with real-world events in verifiable ways. The results highlight how blurred the line is between opinion and fact in social media. Although the dataset included several categories, the final analysis focused on six key themes. Lower-frequency ones like Environmental Concerns and Media Criticism were grouped under Other in the table.

A breakdown of claims by theme revealed further insights. Human Rights and Geopolitics were the most common categories and had the highest proportion of Likely True outcomes, often supported by sources like Human Rights Watch, Amnesty International, or reputable media. Geopolitical and Financial Ethics claims showed a more varied distribution, with more Opinion, Unverifiable, and Likely False verdicts—reflecting the interpretive nature of such topics. Sportswashing also featured heavily and had a notable presence of Likely True claims, reinforcing its central role in public discourse around Gulf involvement in sport.

Category	Likely True	Likely False	Opinion	Unverifiable
Human Rights	4,181	893	3,131	955
Geopolitics	1,884	877	3,327	799
Other	1,149	391	1,696	655
Financial Ethics	1,079	264	1,511	449
Corruption	606	219	743	289
Sportswashing	339	77	508	91

Table 6: Claim Verification Outcomes by Category

Although not all claims could be independently verified, the inclusion of source links and model-generated explanations improved transparency and interpretability. For instance, claims referencing human rights violations were often substantiated with publicly available reports. Two representative examples are shown below, with a wider selection available in Appendix C.

Comment	Verdict	Source(s)
“This exploitation of workers is unfortunately a hallmark of many Gulf countries.”	Likely True	Human Rights Watch
“Millions of people are killed in the Middle East, and no one bats an eye.”	Likely False	BBC, UN Reports

Geopolitical claims showed the most polarity, with some flagged as Likely True and others as Likely False despite referencing similar issues. This highlights how political framing and narrative bias shape public perceptions and how facts are selectively interpreted depending on the commenter’s viewpoint. Comments classified as Likely False or Unverifiable also tended to coincide with strong emotional tones — particularly anger, disapproval, or sarcasm. Somewhat unexpectedly, many of these appeared in negative comments criticising Gulf states rather than in supportive ones, suggesting that misinformation can also arise from oppositional discourse where critics rely on overstated or unverifiable claims. This reinforces existing research showing that misinformation is more likely to appear in emotionally charged discourse. In sum, the claim detection and verification process revealed that YouTube commenters often engage with real-world issues through both fact and opinion. Many claims were verifiable using external sources, and their distribution across themes suggests a meaningful pattern of scrutiny toward Gulf state involvement in sport. While large language models made this large-scale analysis feasible, the subjective tone and political complexity of many claims highlight the importance of human judgment in evaluating their credibility.

C. Transcript Analysis & Results

Due to GPT-4o’s token limits, I compared each comment only to the top 5–8 most relevant transcript sentences using SBERT. This reduced noise but may have led to misclassifying some comments as neutral due to limited context. As a result, the neutral count might be slightly inflated, though alignment results remain accurate within these constraints.

From the 72,000 comments analysed, 56,390 were influenced by the transcript. Of those, 19,832 disagreed, 25,583 were neutral, and 10,974 agreed with the video’s messaging. Disagreement was most common in videos tied to Saudi Arabia and Qatar, particularly where transcripts downplayed or deflected criticism related to human rights, corruption, or foreign influence. This suggests viewers actively pushed back against attempts to soften controversial issues, reacting critically to biased or selective framing. Agreement was more frequent when videos focused on sporting success or infrastructure achievements — especially in content related to Qatar’s World Cup coverage or UAE-based club ownership, where political framing was less prominent. Sentiment patterns highlighted clear contrasts:

- **Disagreeing comments** leaned heavily negative, often expressing anger, disapproval, or frustration. Many flagged hypocrisy, double standards, or attempts to distract from systemic issues.
- **Neutral comments** were the most diverse in tone — some reflected confusion or scepticism, while others maintained a critical edge but lacked clear alignment. Many hinted at agreement or disagreement but didn’t commit clearly, often because the transcript was vague or avoided taking a stance.
- **Agreeing comments** had the highest proportion of positive sentiment relative to size. Often praised the host country’s ambitions, infrastructure, or hospitality, especially under more promotional videos.

The pattern extended into factual claims. Disagreeing comments contained the highest number of check-worthy claims (8,288), of which 2,018 (24.3%) were classified as either Likely False or Unverifiable. These often included emotionally charged assertions about worker conditions, media manipulation, or broader political agendas. While many reflected genuine concerns, others leaned into exaggeration or lacked reliable sourcing. In contrast, agreeing comments produced fewer claims overall, and the ones that were flagged were more likely to be supported by verifiable evidence or grounded opinion.

As a point of comparison, I also extracted and verified factual claims made directly by the transcripts themselves. These claims tended to be more polished and less extreme in tone, often focusing on infrastructure success, cultural pride, or general optimism. However, in several cases, viewer disagreement clustered around transcript claims that were later verified as Likely False or Unverifiable. This suggests that viewers were not just reacting emotionally, but in some cases responding to specific narrative inconsistencies or factual inaccuracies in the video content. It also reinforces the idea that misinformation and resistance often coalesce around how the original narrative is framed.

- **Disagreeing comments:** Marked by emotionally intense reactions, more likely to challenge the narrative’s legitimacy and contain misinformation or unverified claims.

- **Neutral comments:** Varied in tone and content, often reflecting either uncertainty or disengagement due to vague or PR-heavy video content
- **Agreeing comments:** Smaller in volume, but more consistent in tone, with higher factual alignment and clearer expressions of support or approval

These patterns suggest that alignment with the transcript wasn't just shaped by a viewer's stance — it was also triggered by how the content was framed. When videos addressed criticism with transparency, even if not fully convincing, disagreement was less intense. But when the transcript felt evasive, dismissive, or overly celebratory, it created space for pushback. That pushback was often emotional and more likely to involve claims that blurred the line between opinion and fact. In this way, viewer responses reveal more than just sentiment — they reflect the audience's ability to recognize and react to narrative positioning. Where the video invited reflection, viewers engaged. Where it appeared to avoid accountability, resistance became louder — and more polarized.

D. Temporal Trends in Sentiment & Claims

To explore how sentiment and claims shifted over time, I split comments into Pre, During, and Post phases based on each event's timeline. Only events with clear date windows were used. Older cases like Man City and PSG are included in the table for completeness but won't be explored in much detail, since almost all their comments came after the fact — making real-time analysis impossible.

Event	Pre	During	Post
FIFA World Cup 2022	13,180	30,626	5,958
Saudi Pro League	1,710	1,804	2,251
Manchester City Ownership	0	0	4,757
Newcastle United Ownership	1,292	1,006	2,328
LIV Golf	30	1,015	2,386
PSG Ownership	0	0	460

Table 7: Temporal Distribution of Comments by Event Phase

Qatar World Cup 2022

- **Pre-event:** Negative sentiment dominated, especially around human rights violations and migrant labour. Many check-worthy claims appeared here, often unverifiable or exaggerated. Disagreement with the video transcript peaked in this phase.
- **During:** Sentiment shifted toward the neutral zone. Some critical voices persisted, but hospitality, crowd energy, and infrastructure quality softened the tone. The transcript's focus on logistics and spectacle influenced this response.
- **Post-event:** Positive sentiment rose sharply, especially around hospitality and execution. Misinformation dropped, and most claims shifted toward opinion rather than factual challenge.

LIV Golf

- **Pre-event:** Practically no discussion.
- **During:** Comments started to emerge, focused on confusion around the format and motives.
- **Post-event:** Heavy surge in comments. Most were negative, accusing the project of being PR-driven or morally compromised. Claims flagged as Likely False or Unverifiable cantered on sportswashing and financial ethics.

Saudi Pro League

- **Pre-event:** Mostly speculative talk around player transfers and Gulf ambitions. Tone was mixed.

- **During:** Enthusiasm grew slightly, but many comments expressed discomfort with the league's rapid rise and backing.
- **Post-event:** Negative sentiment climbed again, particularly as media hype clashed with perceived lack of authenticity. Claims often tied back to reputation laundering or geopolitical strategy.

What stood out most was how differently each event unfolded over time. The World Cup showed the clearest shift — people started critical and ended mostly impressed. With the Saudi Pro League, it never fully settled. Tone hovered between curiosity and distrust, and even post-event, people weren't convinced. LIV Golf felt like a delayed reaction — barely anything before launch, then a sharp spike in negativity once people paid attention. In contrast, Newcastle's response was steady throughout, suggesting reactions were shaped early and stayed that way.

The themes shifted too. Human rights, corruption, and soft power were prominent early but faded as events progressed. Not because those issues lost relevance, but because the videos changed focus, and the comments followed. Anger was more common before the events, while disapproval became more typical afterward. This shift wasn't just about tone; it reflected a broader redirection of attention. The fading of critical themes may suggest a kind of narrative control. It doesn't necessarily mean sportswashing changed opinions, but it may have limited how long criticism stayed in focus. Intentional or not, this shows how media framing can shape public debate.

E. Network Analysis of Sentiment & Claim Narratives

The graph below visualizes one of the largest comment communities, based on a Qatar 2022 FIFA World Cup video and filtered by sentiment. As shown in the broader sentiment analysis, negative responses dominate, with smaller pockets of neutral and positive reactions. Node size reflects engagement, calculated using a weighted score of likes and replies, meaning larger nodes represent comments that attracted more visibility. These high-engagement nodes tend to cluster around emotionally charged content, particularly those expressing disapproval, frustration, or moral critique.

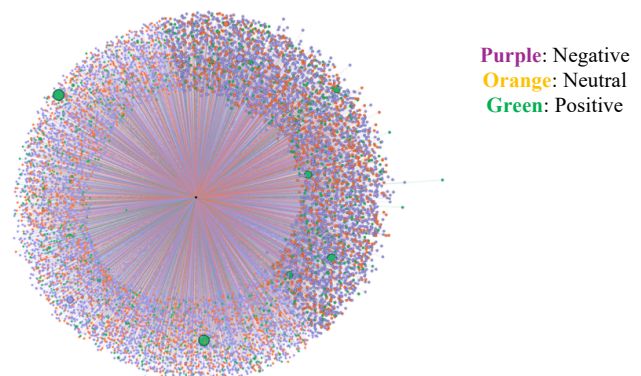


Figure 1: Network of Comments by Sentiment and Visibility

This visual reinforces how public discourse around Gulf sporting events is often polarized and shaped by narrative framing. While some positive sentiment exists, critical perspectives clearly dominate in both volume and visibility. The network's disconnected structure also reflects the platform's siloed design, where discussions remain confined to individual videos, limiting broader narrative interaction or correction. Though this graph focuses on a single high-

engagement community for clarity, additional filtered networks from other events and countries are included in Appendix D, showcasing how sentiment, misinformation, and narrative themes vary across the wider dataset.

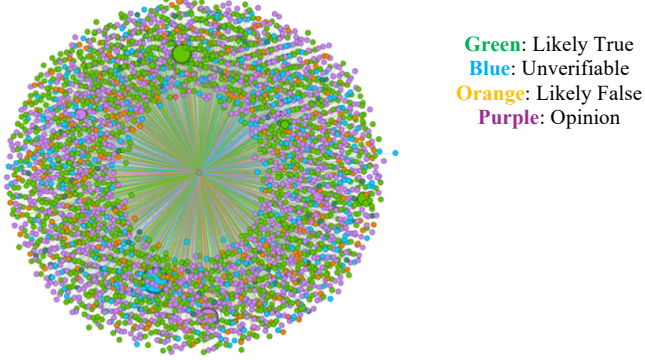


Figure 2: Network of Comments by Claim Verdict

Extending the analysis beyond sentiment, claim verification was mapped onto the same network. This showed how claim credibility shaped the structure and density of discourse. Comments marked Likely False or Unverifiable clustered in emotionally reactive areas, especially around corruption, labor rights, and geopolitical legitimacy. These dense formations suggest that ambiguous or misleading content concentrates where sentiment is most polarized. In contrast, Likely True claims were more diffusely distributed, often spanning larger clusters with mixed sentiment. Their presence suggests a broader, more stable form of engagement, with verifiable information supporting more balanced discourse. Integrating verified outcomes into the network shows how factual, speculative, and false narratives coexist and compete within tightly structured communities. As with sentiment, visibility and engagement appear more tied to emotional intensity than factual accuracy.

F. Sentiment Forecasting: A Look Ahead to 2034

To evaluate whether sentiment around Gulf-sponsored events followed identifiable and generalizable patterns, I fine-tuned a DistilBERT sentiment classification model on a labelled dataset of over 72,000 YouTube comments. The model was trained to distinguish between negative, neutral, and positive sentiment across discourse tied to events such as Qatar 2022, LIV Golf, and the Saudi Pro League. The model achieved a final accuracy of 79.5% and a macro F1 score of 0.76, demonstrating strong overall generalizability. Performance across classes was well balanced, with an F1 score of 0.86 for negative sentiment, 0.68 for neutral, and 0.75 for positive, as shown in Table 8.

Event	Precision	Recall	F1-Score	Support
Negative	0.83	0.89	0.86	8,317
Neutral	0.72	0.64	0.68	4,219
Positive	0.79	0.71	0.75	1,753
Overall	0.78	0.75	0.76	14,289

Table 8: Sentiment Classifier Performance Metrics

These results validate the presence of distinct sentiment signals in comments, showing that attitudes toward Gulf mega-events aren’t arbitrary but follow consistent linguistic and emotional patterns. The high recall and precision for

negative sentiment reflect the stable ways critique and scepticism are expressed across events and platforms. Building on this, the model provides a foundation for anticipating how sentiment may emerge in future Gulf-hosted events—particularly the 2034 FIFA World Cup, expected in Saudi Arabia. Although no predictive labels were applied to future content, the consistent language features across past Saudi-linked investments—like the Saudi Pro League, LIV Golf, and the Newcastle United acquisition—suggest similar discourse will likely appear.

Consequently, the model could be extended in future work to evaluate:

- Simulated future discourse, such as news comment sections or pre-event narratives
- Cross-event transferability, by testing how well the model generalizes to unseen but thematically aligned events

Thus, while not applied in real-time forecasting, this predictive experiment reinforces the thesis’s core argument: that sentiment toward Gulf investments in sport—particularly controversial ones—is not only analysable but also, to some extent, predictable. This underscores the potential for machine learning tools to assist policymakers, media analysts, and event organizers in understanding and pre-empting public opinion trajectories.

V. CONCLUSION & FUTURE WORK

This thesis examined public sentiment and misinformation around Gulf state investments in sport, analysing 72,000 YouTube comments from events like the FIFA World Cup and LIV Golf. Using NLP, large language models, and verification tools, it mapped emotional tone, factual accuracy, and narrative alignment. Results showed dominant negative sentiment—especially toward Saudi Arabia and human rights—with variation across time, countries, and events. Disagreement often clustered around evasive or celebratory narratives, particularly when inconsistencies were perceived. Misinformation appeared most in emotionally charged comments, reinforcing links between sentiment, framing, and belief accuracy. A predictive model trained on past patterns showed responses to Gulf-sponsored events follow recognizable trends. While LLMs enabled large-scale analysis, they also carry risks of bias, hallucination, and misinterpretation. This study offers a scalable framework for analysing how online discourse reflects and contests sportswashing—and supports more transparent, human-aware approaches to digital opinion analysis.

Future research could build on this study by incorporating human annotation to validate LLM outputs and reduce reliance on automated labels. Expanding the analysis across other platforms like Reddit or X could reveal whether sentiment patterns generalize beyond YouTube. Further work might also explore multilingual or cross-cultural differences in sportswashing discourse or apply the predictive model to simulate reactions to upcoming events such as the 2034 World Cup. Finally, incorporating temporal dynamics and user-level interactions into network models could offer deeper insight into how polarized narratives and misinformation emerge, spread, and diminish across media cycles.

REFERENCES

- Siersdorfer, S., Chelaru, S., Nejd, W. and San Pedro, J.,** 2010. How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In: **Proceedings of the 19th International Conference on World Wide Web (WWW '10)**, Raleigh, NC, USA, 26–30 April 2010. New York: ACM, pp. 891–900. doi:10.1145/1772690.1772781.
- Suter, V., Shahrezaye, M. and Meckel, M.,** 2022. COVID-19 Induced Misinformation on YouTube: An Analysis of User Commentary. *Frontiers in Political Science*, 4:849763. doi:10.3389/fpos.2022.849763.
- Ganji, S.K.,** 2023. The rise of sportswashing. *Journal of Democracy*, 34(2), pp.75–89. Available at: <https://www.journalofdemocracy.org/articles/the-rise-of-sportswashing> [Accessed 27 July 2025].
- Grix, J., Brannagan, P.M., Grimes, H. and Neville, R.,** 2023. Unpacking the politics of ‘sportswashing’: It takes two to tango. *Political Studies Review*, [online] Available at: <https://doi.org/10.1177/14789299231214546> [Accessed 27 July 2025].
- Gerschewski, J., Giebler, H., Hellmeier, S., Keremoğlu, E. and Zürn, M.,** 2024. The limits of sportswashing. How the 2022 FIFA World Cup affected attitudes about Qatar. *PLOS ONE*, 19(8), e0308702. doi:10.1371/journal.pone.0308702.
- Aly, R., Guo, Z., Schlichtkrull, M., Thorne, J., Vlachos, A., Christodoulopoulos, C., Cocarascu, O. and Mittal, A.,** 2021. FEVEROUS: Fact Extraction and VERification Over Unstructured and Structured information. In: *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Track on Datasets and Benchmarks*. arXiv:2106.05707 [cs.CL]. Available at: <https://arxiv.org/abs/2106.05707> [Accessed 27 July 2025].
- Schlichtkrull, M., Ousidhoum, N. and Vlachos, A.,** 2023. The intended uses of automated fact-checking artefacts: Why, how and who. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, pp.8618–8642. Association for Computational Linguistics. Available at: <https://aclanthology.org/2023.emnlp-main.591> [Accessed 27 July 2025].
- Islam T., Zaman S.M.M., Jain V., Rani A., Rawte V., Chadha A. and Das A.,** 2024. A comprehensive survey of hallucination mitigation techniques in large language models. arXiv preprint arXiv:2401.01313 [cs.CL]. Available at: <https://doi.org/10.48550/arXiv.2401.01313> [Accessed 27 July 2025].
- Brandwatch,** 2022. What is a good engagement rate on social media? Brandwatch blog, 14 January. Available at: <https://www.brandwatch.com/blog/what-is-good-engagement-rate-on-social-media/> [Accessed 27 July 2025].
- Thorne, J., Vlachos, A., Christodoulopoulos, C. and Mittal, A.,** 2018. FEVER: A large-scale dataset for fact extraction and verification. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, pp.809–819. Available at: <https://aclanthology.org/N18-1074/> [Accessed 20 Aug. 2025].

Rewritten Comment

f"""

You are a text cleaner. Your task is to rewrite the following YouTube comment to make it more readable while keeping its original meaning intact.

- Fix typos and grammatical errors.
- Expand abbreviations where necessary.
- Keep slang words if they contribute to the meaning but clarify if needed.
- **Do NOT** remove or censor foul language.
- Do NOT change the sentiment or tone of the comment.

Original Comment:

"{comment}"

Rewritten Version:

"""

Thesis Relevant

Definition of Sportswashing:

- When sports are used to improve a country's reputation while hiding **human rights abuses, corruption, or political issues**.




- Example: **Gulf states** (Saudi Arabia, Qatar, UAE) investing in sports, hosting events (FIFA, F1), or owning clubs (Man City, PSG, Newcastle).

Classification Rules:

- **'YES'**: Mentions sportswashing, Gulf investments, corruption, political influence, financial takeovers, or criticism/support of Gulf involvement.

- **'MAYBE'**: Unclear connection but mentions Middle Eastern entities, Gulf countries, or potential geopolitical influence.

- **'DEFINITELY NOT'**: Clearly about **match performance, goals, players, or unrelated topics** (e.g., "That was a great goal!", "This team played well").

Emoji & Indirect Mentions: If the comment includes **emojis** (, , ) referring to a Gulf country, **lean** towards 'YES'.

YouTube Comment (Rewritten for Clarity):

"{comment}"

Instructions:

- Respond **ONLY** with **'YES'**, **'MAYBE'**, or **'DEFINITELY NOT'**, nothing else.

- If uncertain, **lean** toward **'YES'** if the comment references any political, financial, or ethical aspect of sports.

"""

Comment Category

f"""Comment: "{comment}"\n\nWhat category does this comment fall under?\n\nReply **ONLY** with one label:\n- Sportswashing\n- Human Rights\n- Financial Ethics\n- Corruption\n- Geopolitics\n- Media Criticism\n- Other"""

Country

f"Below is a YouTube comment and the related sports event. Identify which Gulf state (e.g., Saudi Arabia, Qatar, UAE) is most associated with this comment. Use the event context to help you.

Comment: "{comment}"

Event: "{event}"

If no Gulf state is clearly implied, respond with an empty string. Otherwise, respond with just one country name: Saudi Arabia, Qatar, or UAE."

Event Type

f"Below is a YouTube comment. Your task is to identify which, if any, of the listed events the comment is referring to.

Comment: "{comment}"

Choose only **ONE** of the following events. If the comment does not relate to any, reply with an empty string:

- Manchester City ownership
- Paris Saint-Germain ownership

<ul style="list-style-type: none"> - Newcastle United ownership - FIFA World Cup 2022 - Formula 1 - Saudi Pro League - LIV Golf - Gulf multi-sport hosting"
Sentiment
f""Comment: "{comment}"\n\nHow would you classify the sentiment of this comment?\n\nReply ONLY with:\n1 = positive\n0 = neutral\n-1 = negative""
Emotion Class
f""Comment: "{comment}"\n\nWhat is the dominant emotion expressed in this comment?\n\nReply ONLY with one of the following:\n- anger\n- sadness\n- joy\n- hope\n- disapproval\n- neutral""
Fact Or Opinion
f""Comment: "{comment}"\n\nDoes this comment state a fact or express an opinion?\n\nReply ONLY with:\n1 = factual\n0 = opinion""
Claim Detection
f""Comment: "{comment}"\n\nDoes this comment contain a verifiable factual claim?\n\nReply ONLY with:\n1 = yes\n0 = no\n\nReply with just the number: 1 or 0""
Influenced By Transcript
f""Comment: "{rewritten_comment}" Transcript: <ul style="list-style-type: none"> - {top_sentences[0]} - {top_sentences[1]} - {top_sentences[2]} - {top_sentences[3]} - {top_sentences[4]} Was this comment influenced by the transcript? Reply ONLY with: 1 = Yes, influenced 0 = No, not influenced""
Agreed With Transcript
f""Comment: "{comment}"\n\n{fallback}\n\nDoes this comment agrees with the transcript?\n\nReply ONLY with:\n1 = agrees\n-1 = disagrees\n0 = neutral""

APPENDIX B

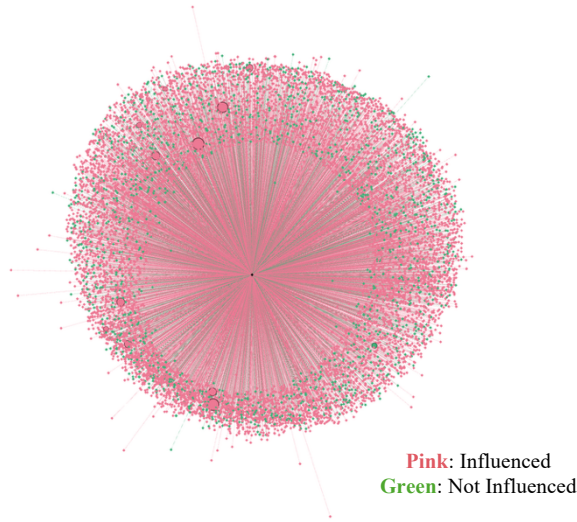
MISCLASSIFICATION EXAMPLES

Comment	Why it might be misclassified
"At 0:09, football is sponsored by 'CORONA.' LOL."	Labelled as Likely True but includes informal language ("LOL") and could be sarcastic. Explanation references the sponsorship, but the tone of the claim makes factual certainty questionable.
"Venezuela has more oil than Saudi Arabia. That is a fact. Just imagine what they could accomplish if it weren't run by leftist communists!"	While the first sentence is factual, the second introduces speculation and strong political opinion. The overall claim blends verifiable information with subjective judgment, which may confuse a language model into labelling it fully factual.
"Basically, if Sam Smith went to Saudi Arabia wearing nipple tassels and hot pants with a wedgie in public, unlike in the UK, he would be arrested. Just don't be weird, and you won't get arrested. 😊"	The claim is speculative and hypothetical, phrased with informal language and a sarcastic tone. Despite being labelled as likely true, it reflects personal judgment rather than a verifiable statement.
"This is the same guy who spent half a billion dollars on a painting that sits on his super yacht."	The claim has a mocking, informal tone and lacks direct evidence. It blends partial fact with rhetorical framing, making it difficult for a model to evaluate strictly on factual grounds.
"This is where the old saying goes, 'Power corrupts, and absolute power corrupts absolutely.'"	The quote is proverbial, not a factual claim. Labelling it as likely true treats moral commentary as verifiable fact.

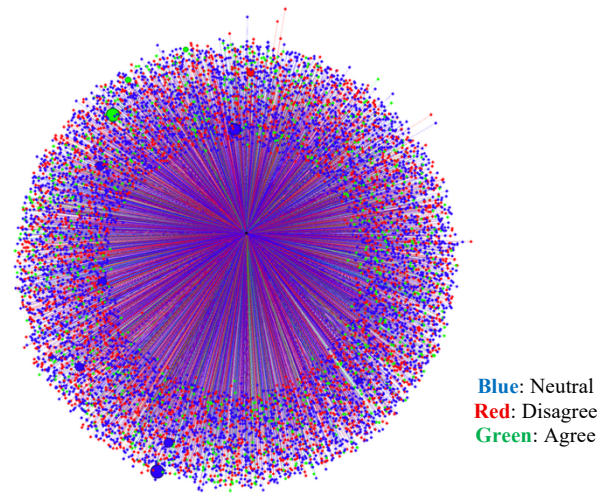
APPENDIX C

CLAIM EXAMPLES

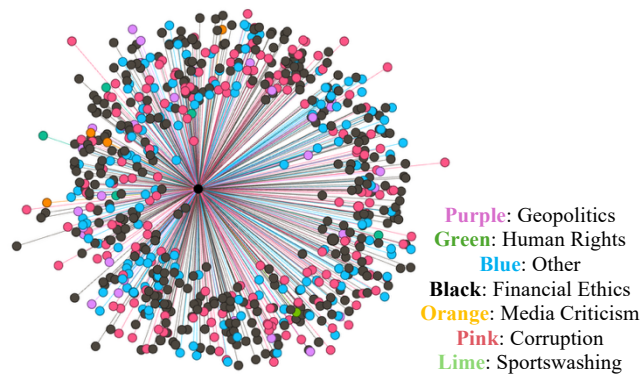
Claim	Verdict	Source(s)
"I have also realized that the LGBTQ community is now vibrant in the game. The captain's armband and corner flags display the LGBTQ flag."	Likely True	Premier League, BBC Sport
"Basically, if Sam Smith went to Saudi Arabia wearing nipple tassels and hot pants with a wedgie in public, unlike in the UK, he would be arrested. Just don't be weird, and you won't get arrested. 😊"	Likely True	CNN, The National, Lonely Planet
"Fact: In Islam, women have been allowed to ask for a divorce for 1,400 years."	Likely True	Islami City, Al-Islam.org, BBC
"As you are discussing freedom of speech, it is non-existent in Myanmar and in Bangladesh as well."	Likely False	Human rights watch, amnesty international
"Alcohol is poison, and women don't have to worry; all the rapists are in Europe."	Likely False	Mayo Clinic, UN Women
I thought he couldn't compete in the U.S. because he has one eye. 😊	Likely False	Disabled Sports USA, Paralympic.org
"Olympics in Rio... no condoms means a higher risk of spreading HIV. To athletes... the red light district is also not a good example."	Opinion Or Speculation	No reliable source
"How is suppressing gay rights a crime? It's a Muslim country, with the holiest place in Islam located there: the Kaaba in Mecca, which we face while praying five times a day. There is no place for homosexuality in Islam, just as there isn't in authentic Christianity or Judaism. Shunning and getting rid of, as well as correcting, homosexuals is not a crime. It never has been. Being homosexual goes against the very laws of nature and morality."	Opinion Or Speculation	No reliable source
"You, Piers, are a racist who supports white supremacy and promotes the theft of land from other people. You advocate for the killing and murdering of innocent civilians. Stop pretending."	Opinion Or Speculation	No reliable source
One thing is for sure: whether single or not, women will feel very safe in Saudi Arabia. There's no chance of getting harassed or receiving dirty looks. Men are very respectful towards women. I love India, but compared to Saudi Arabia, tourist women are 1,000 times safer there, whether it's day or night, than in India. Don't just take my word for it—look at the statistics.	Opinion Or Speculation	No reliable source



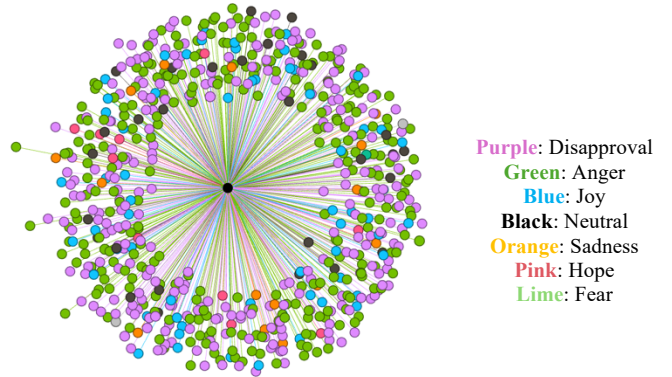
Qatar, 2022 World Cup (Influenced by Transcript)



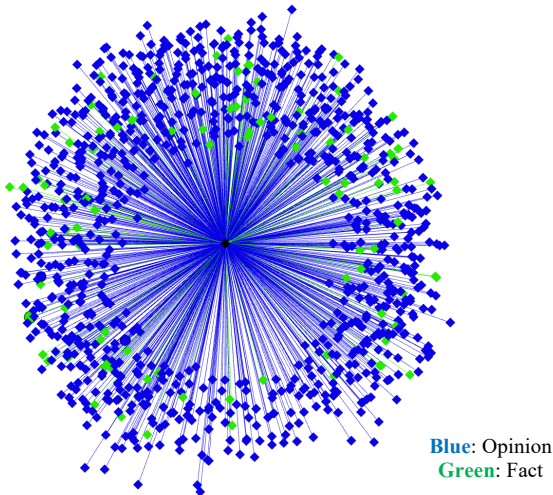
Qatar, 2022 World Cup (Agreed with Transcript)



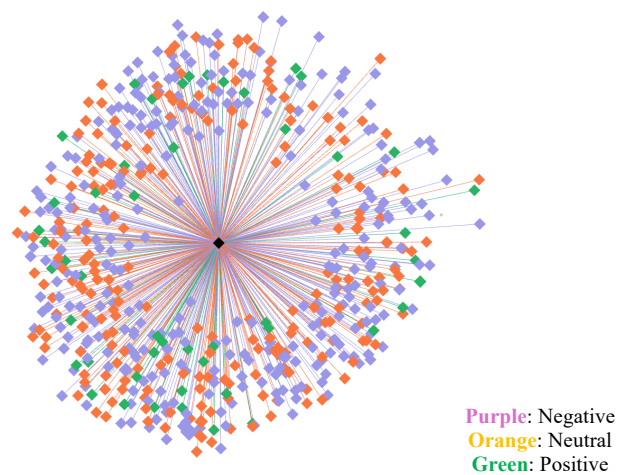
UAE, Man City Ownership (Category)



Saudi Arabia, Newcastle Takeover (Emotion Class)



Saudi Arabia, LIV GOLF (Fact or Opinion)



Saudi Arabia, Saudi Pro League (Sentiment)

Note: These Gephi network visualisations depict relationships between YouTube comments and the videos they belong to. Nodes represent either comments or videos, with node sizes scaled by Engagement Score (larger nodes indicate higher engagement). Edges represent the connection between a comment and its associated video. Each visual is coloured according to a specific analytical attribute shown in its legend (e.g., influence status, agreement level, category, factuality, emotion, sentiment), enabling comparison across different dimensions of the analysis.