THREAT MATRIX

Signify | respondology

FIFA / FIFPRO

SOCIAL MEDIA PROTECTION SERVICE

FIFA WOMEN'S WORLD CUP
AUSTRALIA & NEW ZEALAND 2023™

FULL TOURNAMENT ANALYSIS

Signify | respondology

CONTENTS

THREAT MATRIX
FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

INTRODUCTION

This report represents a summary analysis of all monitoring and moderation activities carried out by FIFA's Social Media Protection Service (SMPS) across FIFA Women's World Cup Australia & New Zealand 2023™.

The data and insights in this document were gathered from Wednesday 19th July through to Monday 21st August 2023, incorporating 64 fixtures of FIFA Women's World Cup 2023™.

The report covers SMPS activities across all major platforms including X (formerly Twitter), Instagram, Facebook, TikTok and YouTube, and is supplementary to the daily reports provided to FIFA across the tournament.

Launched in 2022, FIFA's Social Media Protection Service has covered the FIFA World Cup Qatar 2022™ and five other FIFA affiliated tournaments leading up to the FIFA Women's World Cup 2023™.

The service has delivered the most comprehensive level of defence to players, coaches and officials from social media abuse ever activated.

This report will provide the following analysis:

- Monitoring data + insights
- Moderation activities
- Issue categorisations
- Timelines of abuse
- Targeted national squads
- Account investigations

5.1m
posts / comments analysed

103k
posts / comments flagged by Threat Matrix AI and reviewed by humans

7,085
posts / comments verified as abusive reported to platforms

117k
comments hidden

5.8k
unique accounts detected sending abusive posts / comments

628
verified identities of abusive accounts

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

MONITORING + MODERATION

FIFA introduced the Social Media Protection Service (SMPS) to provide participating teams, players, officials and other stakeholders during FIFA tournaments with the following:

- **Monitoring** – FIFA monitors public accounts and mentions for discriminatory, abusive or threatening content. No action is required from individuals or representatives for this service to run. Where content is detected and verified FIFA takes action working in partnership with platforms for swift removal.

- **Moderation** – an individual or representative can opt-in to moderation of their account which allows abusive content to be detected and hidden in real time.

The central goal of these services is to protect players, teams, officials and fans from abuse, keeping their social feeds free from hate and allowing them to focus on enjoying their part in FIFA events.

Players, teams and officials are protected against discriminatory, abusive or threatening content in FIFA's official languages and the official languages of participating teams.
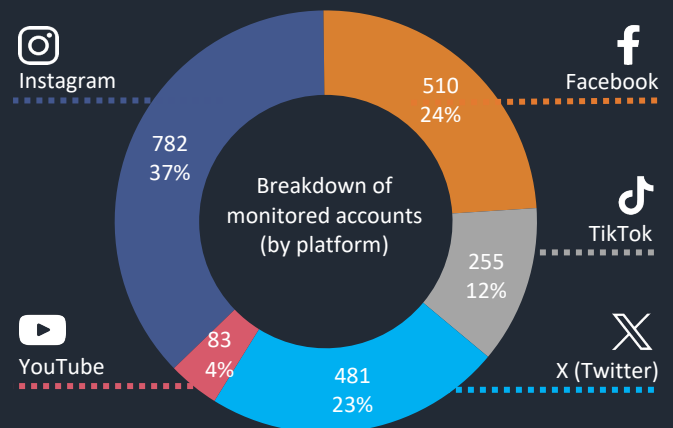
ACCOUNTS MONITORED

2,111
active accounts covered across five platforms

During the FIFA Women's World Cup Australia & New Zealand 2023™ protective coverage was set up for the following participants:

- 697 Players and Coaches with 1,805 active accounts
- 29 Officials with 37 active accounts
- 32 Teams with 202 active accounts
- 35 Ex-Players / Media with 63 active accounts
- 4 active tournament accounts

Abusive posts / comments are identified using text, emoji and phrase categorisation. An AI-empowered threat detection algorithm is also applied.

Breakdown of monitored accounts (by platform)

Instagram 782 37%
Facebook 510 24%
TikTok 255 12%
X (Twitter) 481 23%
YouTube 83 4%

A BESPOKE PROCESS: BUILT FOR FOOTBALL

Build a proactive net around players: AI scan for abuse of player's accounts across social media platforms.

Instantly hide abusive comments in real time on opted-in team and player accounts.

Proactively identify accounts used by abusers at scale - using a combination of AI tech and the nuance of security and intelligence experts.

Unmask abusers using specialist Open Source Intelligence forensic tools to **de-anonymise and identify abusers to an evidential standard.**

Report abusive accounts to social media platforms, Member Associations and Law Enforcement, removing the expectation on players to report abuse.

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

SMPS COVERAGE SINCE LAUNCH

FIFA's SMPS has been active across eight global competitions, including the FIFA World Cup Qatar 2022™ and the FIFA Women's World Cup Australia & New Zealand 2023™. **The level of player coverage, monitoring, reporting and overall protection is unprecedented in global sport.**

6,019
accounts protected by the SMPS across eight global competitions

2,901
participants protected and 143 teams covered

28,009
posts / comments reported to platform

403,715
comments hidden

27.4m
posts / comments scanned

FIFA WORLD CUP Qatar 2022

FIFA WOMEN'S WORLD CUP 2023™

FIFA U-20 WORLD CUP ARGENTINA 2023

FIFA CLUB WORLD CUP

FIFA CLUB WORLD CUP

FIFA NATIONS CUP

FIFA WORLD CUP

FIFA U-17 WORLD CUP INDONESIA 2023

FIFA CLUB WORLD CUP Saudi Arabia 2023

1,200
1,000
800
600
400
200
-

Nov-Dec 2022 | Feb 2023 | May – Jun 2023 | ← July 2023 → | Jul-Aug 2023 | Nov 2023 | Dec 2023
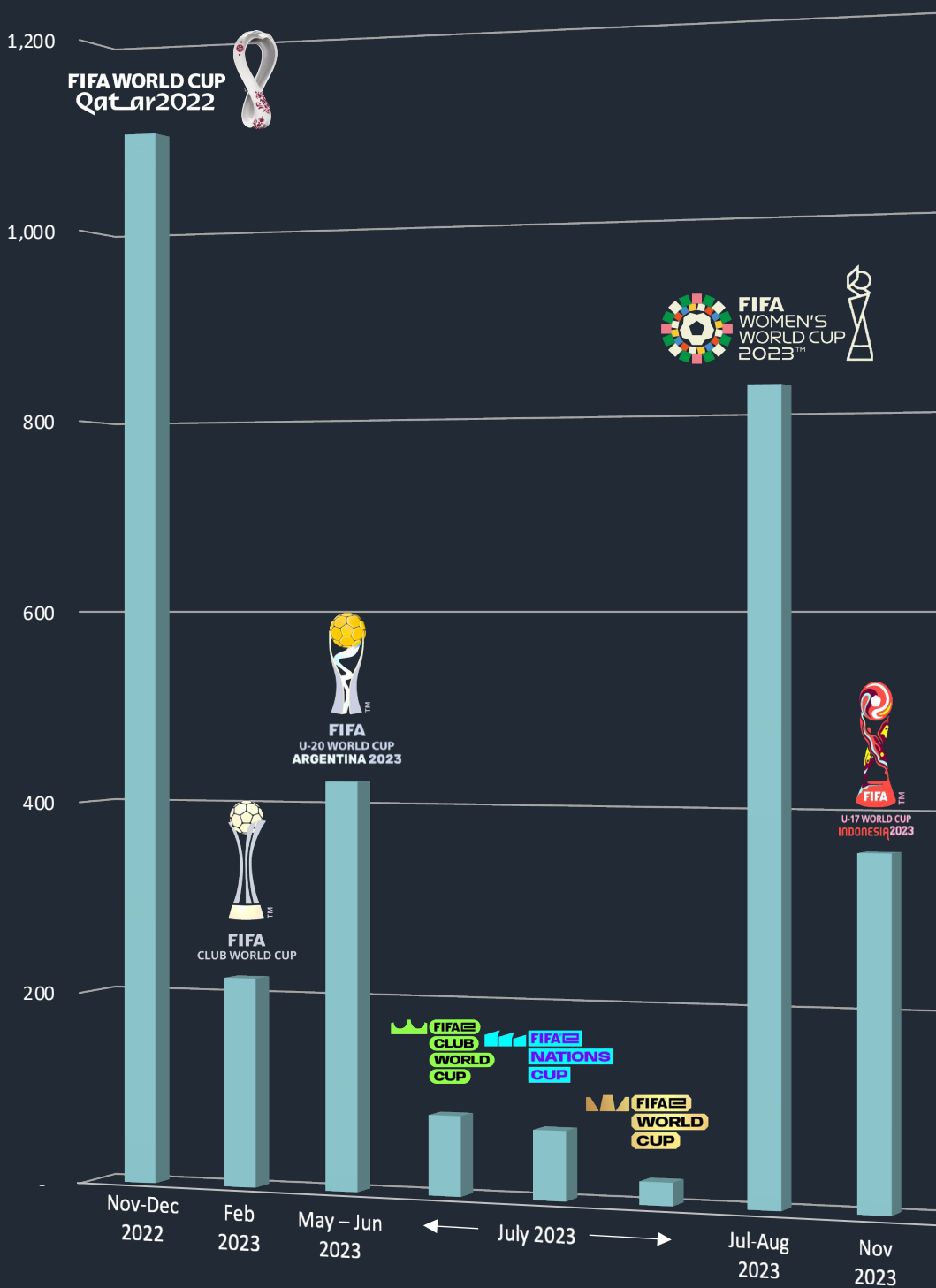
THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

BREAKDOWN OF FINDINGS

FIFA Women's World Cup 2023™ saw discriminatory, abusive or threatening content targeted at over 150 players, but two teams stood out as key targets – the USA and Argentina.

The profile of the USA squad (coming into the tournament as winners of the previous two tournaments), made them a target for online abuse. This was heightened by the perception of players not singing the National Anthem being called out as unpatriotic and anti-American.

One Argentina player was also on the receiving end of large volumes of the detected abuse, whilst politicians of many nations triggered abuse by offering support to their teams, and sometimes commented negatively about performance, producing an increased influx of abusive posts / comments.

SUMMARY ISSUES

1. **US Women's National Team** were the most targeted in the tournament.

2. **Two Players were targeted above all others**. One from USA and one from Argentina.

3. **Politicians'** posts / comments (mostly of support), triggered large sections of abuse.

4. **Threat was prevalent** throughout the tournament, often within sexually abusive posts / comments.

5. **Match Officials** were also targeted directly.

KEY INSIGHTS

▪ **1 in 5 players targeted:** more than 150 players (out of 697 active players monitored) were targeted with some form of discriminatory, abusive or threatening content.

▪ **Homophobia was prolific:** with almost twice the % vs FIFA World Cup Qatar 2022™, platform responses appeared blurred by the cultural differences which seemed to prevent action. **Sexualised content** was also prolific.

▪ **Inconsistent response from platforms:** platforms are dealing better with some incidents but there is still a highly sporadic approach to tackling and removing reported content.
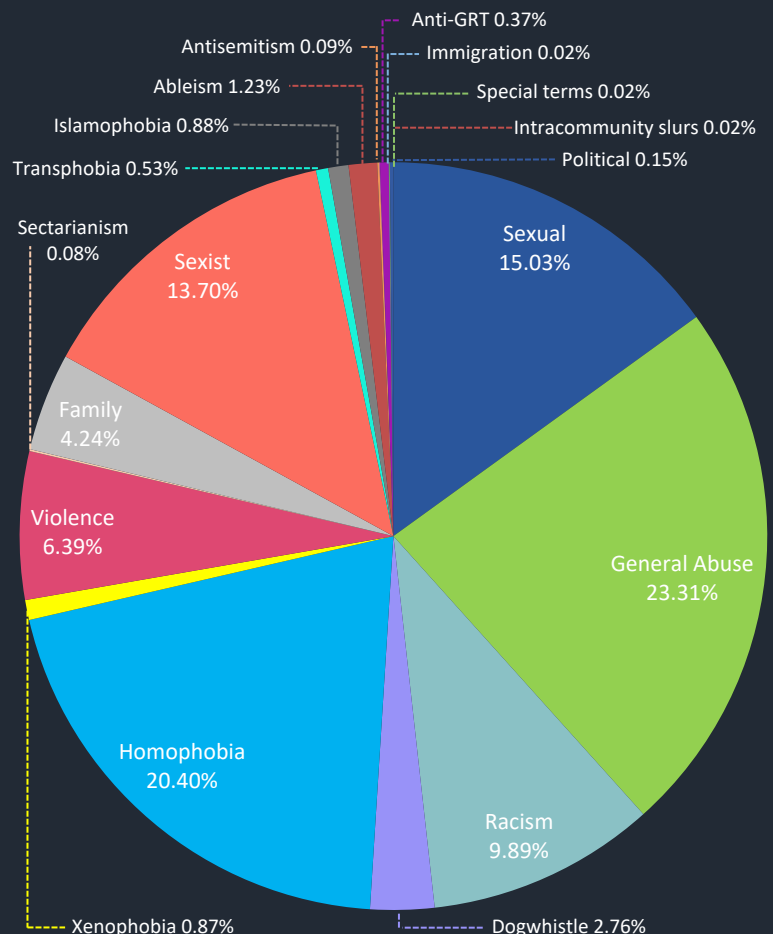
Detected abusive messages by category: all platforms
(19th July – 21st August 2023)

6

THREAT MATRIX
FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

GEOGRAPHIC LOCATION

Verified geography of discriminatory, abusive or threatening accounts (sources of posts / comments).

This chart is a representation of accounts where author location has been verified, across the world, drawn from abuse reported to social media platforms at the FIFA Women's World Cup Australia & New Zealand 2023™.
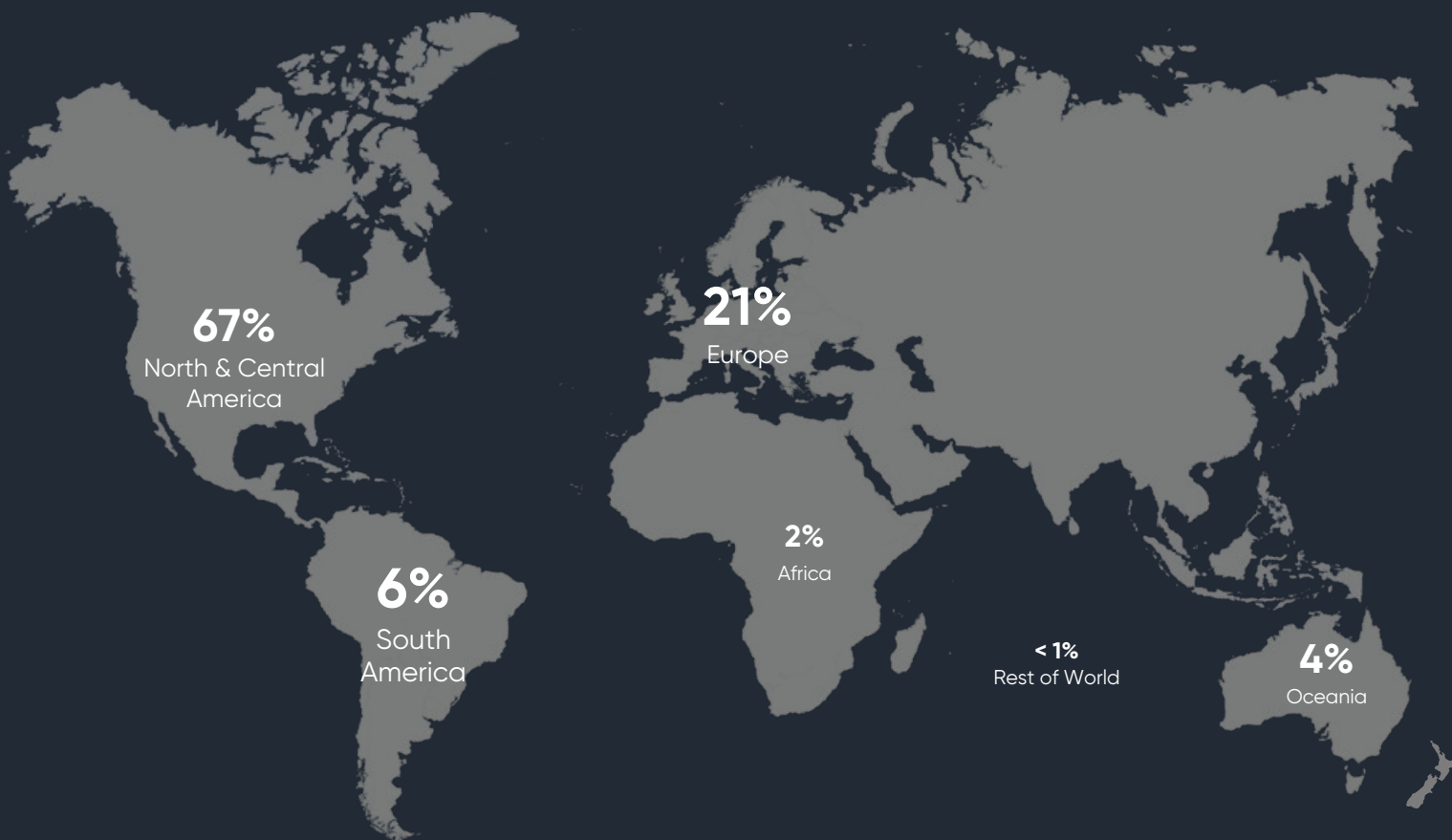
67%
North & Central America

21%
Europe

6%
South America

2%
Africa

< 1%
Rest of World

4%
Oceania

Based on 1,338 social media accounts with identified locations

THREAT MATRIX
FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

MONITORING KEY STATS + INSIGHTS

The SMPS proactive monitoring service has scanned over 5 million posts / comments on X (formerly Twitter), Instagram, Facebook, TikTok and YouTube throughout the tournament.
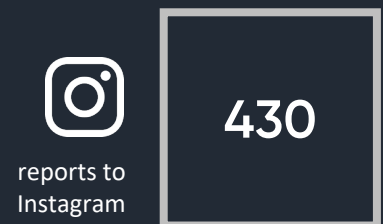
The system flagged 102,511 posts which went through a human review process – of which 7,085 were verified as discriminatory, abusive or threatening and reported to platforms.

Key platform insights:

- X has by far the highest volume of identified, targeted discriminatory, abusive or threatening content, with initial takedown rates of 39%.

- Where there was additional escalation to X after failure to act on immediate automated reporting, a takedown rate of over 95% was recorded - highlighting the value of direct platform engagement. The ability to escalate was not made available by all platforms.

- Meta has committed to reviewing all supplied posts / comments for consideration of takedown post-tournament – reports will be revisited alongside an expected status update from Meta to review activity and action.

- Other platforms continue to perform well with regards to actioning flagged material provided by the SMPS team with TikTok and YouTube both performing consistently with previous events – displayed high percentage takedown rates (80% - 100%).

7,085
verified instances of abuse or threat reported to platforms following Threat Matrix review*

X reports to X (formerly Twitter)	6,184
Instagram reports to Instagram	430
Facebook reports to Facebook	150
TikTok reports to TikTok	321
YouTube reports to YouTube	7

* Abuse totals based on the number of reported breaches of platform's own community guidelines.

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

REAL–WORLD ACTION

These activities are intrinsically linked to real-world actions that start with submission to platforms and follow through to account investigations and detection of the owners of discriminatory, abusive or threatening accounts.

The SMPS Open Source Intelligence team (OSINT) have reviewed and triaged every discriminatory, abusive or threatening message, identifying (where possible), the account owners and gathering any available evidence to take action to prevent further targeted online abuse.

The figures (right) relate to the triage categories applied to accounts sending verified abuse. These are as follows:

- GREEN: the number of identified and verified account owners detected from posts sent.

- AMBER: the number of accounts being reviewed as part of the SMPS initiative. Many of these accounts will continue to be analysed in the months following the event.

- RED: represents accounts that are likely to require a disproportionate level of resource to fully identify. Recommendations for action therefore focus on accounts marked Green and Amber.

Full analysis and evidence packs have been prepared and, where viable, presented to Member Associations and jurisdictional law authorities.*

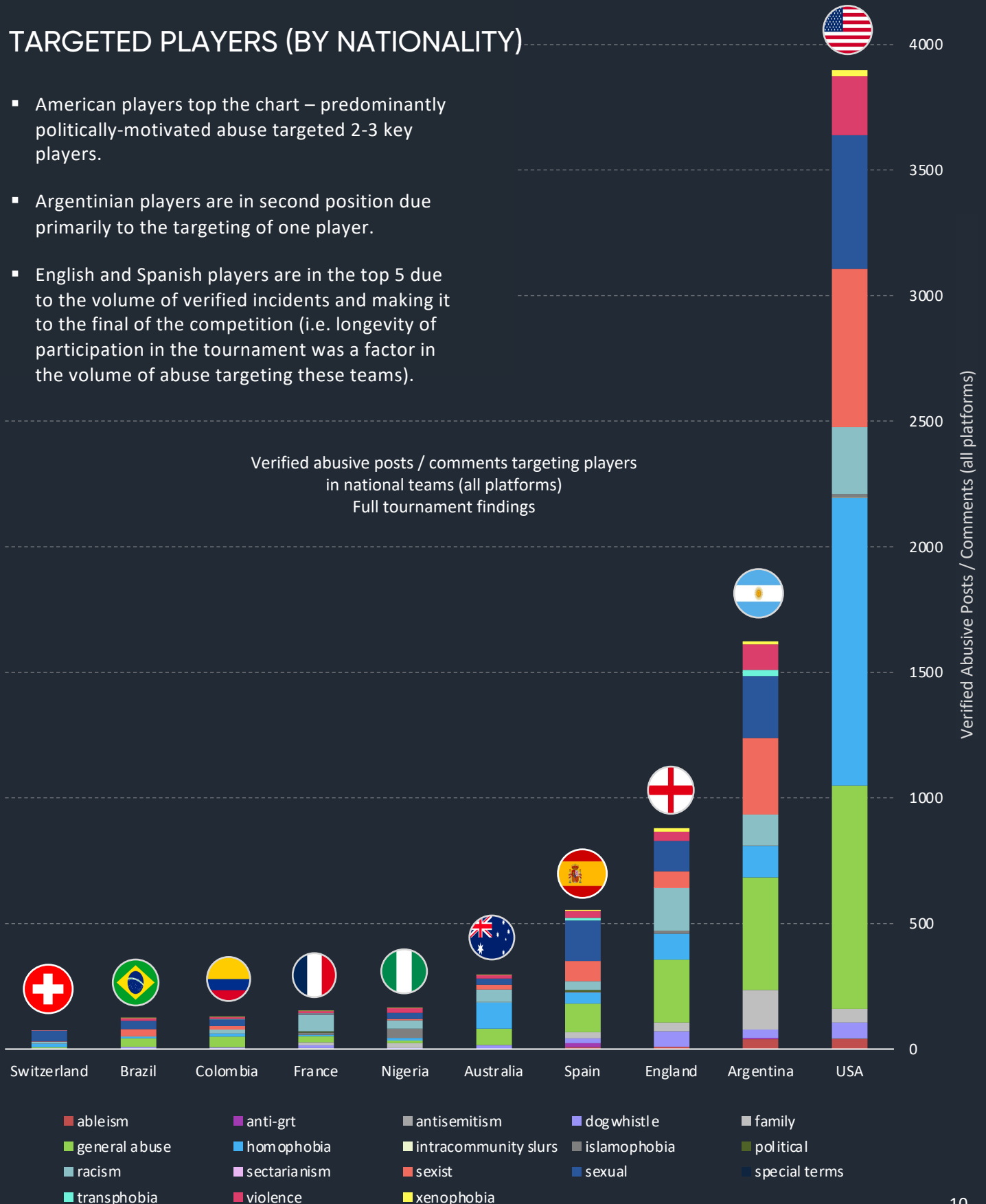| verified identity of abusive account owners | **628** |
| high probability of identification of account owner | **2,007** |
| low probability of identification of account owner or account suspended | **3,570** |

* Where club affiliation has been verified, recommendations to Member Associations can involve contacting the relevant club to apply sanctions / bans. Law Enforcement actions differ by jurisdiction and recommendations for action are tailored accordingly.

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
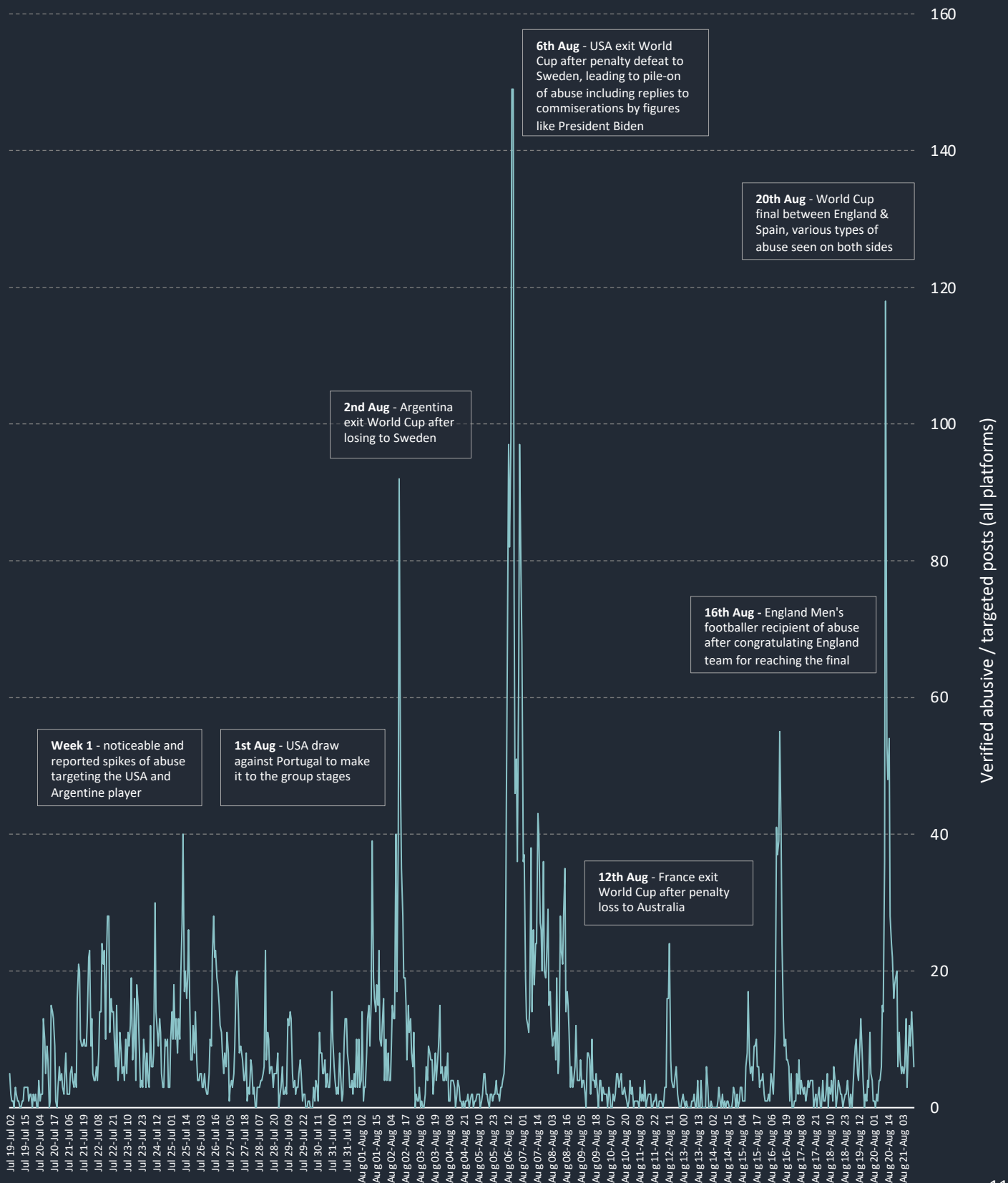### WOMEN'S WORLD CUP 2023™ – ANALYSIS

TARGETED PLAYERS (BY NATIONALITY)

- American players top the chart – predominantly politically-motivated abuse targeted 2-3 key players.

- Argentinian players are in second position due primarily to the targeting of one player.

- English and Spanish players are in the top 5 due to the volume of verified incidents and making it to the final of the competition (i.e. longevity of participation in the tournament was a factor in the volume of abuse targeting these teams).

Verified abusive posts / comments targeting players
in national teams (all platforms)
Full tournament findings

Verified Abusive Posts / Comments (all platforms)

X-axis: Switzerland, Brazil, Colombia, France, Nigeria, Australia, Spain, England, Argentina, USA

Legend:
- ableism
- anti-grt
- antisemitism
- dogwhistle
- family
- general abuse
- homophobia
- intracommunity slurs
- islamophobia
- political
- racism
- sectarianism
- sexist
- sexual
- special terms
- transphobia
- violence
- xenophobia

THREAT MATRIX
FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

DETECTED VERIFIED ABUSE TIMELINE: ALL PLATFORMS

6th Aug - USA exit World Cup after penalty defeat to Sweden, leading to pile-on of abuse including replies to commiserations by figures like President Biden

20th Aug - World Cup final between England & Spain, various types of abuse seen on both sides

2nd Aug - Argentina exit World Cup after losing to Sweden

16th Aug - England Men's footballer recipient of abuse after congratulating England team for reaching the final

Week 1 - noticeable and reported spikes of abuse targeting the USA and Argentine player

1st Aug - USA draw against Portugal to make it to the group stages

12th Aug - France exit World Cup after penalty loss to Australia

Verified abusive / targeted posts (all platforms)

THREAT MATRIX
FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

MODERATION ACTIVITIES

The SMPS has been monitoring and reporting discriminatory, abusive and threatening posts / comments across 30+ languages throughout FIFA Women's World Cup 2023™. Where teams / players have consented and onboarded onto the SMPS moderation service, abusive messages have also been intercepted and instantly hidden.

From 1.3m comments, the service has flagged and hidden over 116k junk, spam, and discriminatory, abusive or threatening posts / comments on Instagram, Facebook and YouTube throughout the tournament.

Comments were hidden on official team channels as well as player accounts.

Levels of abuse can often be driven by frequency of posts and content so some Member Association and player accounts will attract more targeted messages than others.

How does moderation work?

An individual or their representative can opt-in to moderation of their account which allows abusive, discriminatory or threatening content to be detected and hidden in real time.

This service was made available to all players across FIFA Women's World Cup Australia & New Zealand 2023™.

X currently lacks the functionality for a user to fully hide a reply to one of their tweets resulting in this element not being covered by the SMPS moderation and TikTok does not allow automatic, API-driven moderation of comments.

116,820
comments hidden
(player / coach / team accounts)

48,405
moderated comments on Instagram
(player / coach / team accounts)

68,277
moderated comments on Facebook
(player / coach / team accounts)

138
moderated comments on YouTube
(player / coach / team accounts)

9%
average number of reviewed posts / comments hidden by the SMPS service

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

FIFA WOMEN'S WORLD CUP 2023™ FINAL ANALYSIS

The Final of FIFA Women's World Cup 2023™ was a tight match – won 1-0 by Spain over European Champions England.

Key insights and observations detected before, during and after the final included:

BEFORE THE FINAL

- Issues were detected in the build-up to the Final, particularly around the England team. News that members of the UK Royal Family were choosing not to attend the game and the UK Prime Minister wishing the team luck both generated messages containing abusive content (a common trend across this study).

DURING THE FINAL

- As the losing team, England's official accounts received the initial incidents of detected abuse, with specific players singled out for attention.

- The US match official refereeing the final received abuse on X following her involvement in a penalty decision (called by VAR).

AFTER THE FINAL

- Spanish players received targeted abuse, with a more complex set of triggers including congratulatory messages from politicians (another common trigger).

- In the aftermath of the Final, the incident involving Spanish FA President Luis Rubiales and a Spanish player created a significant spike of posts / comments containing abusive (sexual and misogynist) content.

These incidents generated the largest spike of abusive, discriminatory and violent content across the tournament.

On the day of FIFA Women's World Cup 2023™ Final and 24hrs after…

280k
posts / comments captured for analysis

11,612
posts / comments flagged by AI monitoring for human review

637
verified instances of abuse or threat reported to platforms following review*

6,518
comments hidden by SMPS (player / coach / team accounts)

* Abuse totals based on the number of reported breaches of platform's own community guidelines.

THREAT MATRIX
## FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

COMPARATIVE ANALYSIS

Whilst FIFA World Cup Qatar 2022™ was a bigger event in terms of viewership and overall volume of messages mentioning player handles, a relative comparison demonstrates that players at FIFA Women's World Cup Australia & New Zealand 2023™ were subject to a higher percentage of discriminatory, abusive or threatening content.

Players at FIFA Women's World Cup 2023™ were

29%

more likely to be targeted with online abuse than players at FIFA World Cup Qatar 2022™

5.1m
posts / comments mentioning player handles

7k
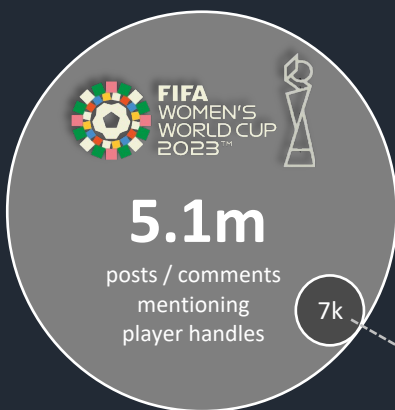
verified / reported abusive posts / comments

19.6k

20m
posts / comments mentioning player handles

FIFA WORLD CUP Qatar 2022

Although a lower raw number, the smaller total engagement around FIFA Women's World Cup 2023™ means that comparatively, players were 28.5% more likely to receive verified abuse or threat.

- **0.14%** of posts / comments captured for analysis during FIFA Women's World Cup 2023™ by the SMPS monitoring system were confirmed to be abusive (7k out of 5.1m).

- **0.10%** of posts / comments captured for analysis during FIFA World Cup Qatar 2022™ by the SMPS monitoring system were confirmed to be abusive (19k out of 20m).

This is significant given the increasing viewership for the FIFA Women's World Cup 2023™, indicating a possible trend as TV / online audiences grow.

Detected segments of abuse were surprisingly similar across both World Cup events, with the exception of Homophobia, which was almost twice as big (as a percentage of overall verified abuse) at FIFA Women's World Cup 2023™.

*visuals are indicative / representative - not to scale.
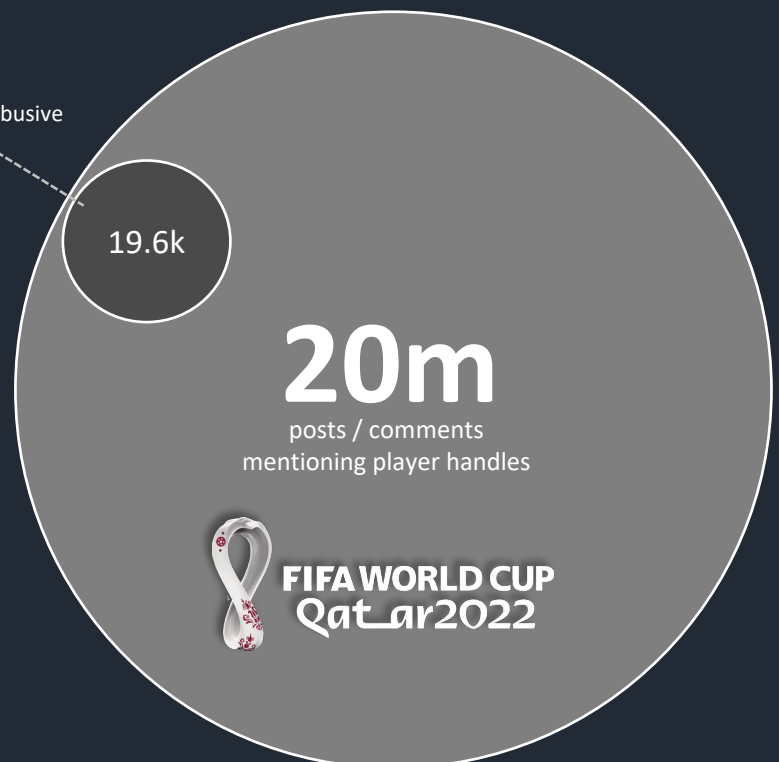
THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

THE MENTAL HEALTH TOLL ON PLAYERS

The hatred and abuse that exists online is a social crisis that touches individuals all over the world and cannot be simply ignored or easily shrugged off. In professional football, this toxic online environment is a difficult and risky place to be in for players.

Footballers are heavily exposed to the continuous vitriol of online trolls and abusers, with high-profile players being the most vulnerable to large-scale abuse, as evidence shows.

The risks and mental health challenges associated with being victim of online abuse are real and have a direct and immediate effect on players. Hatred and discrimination in the online environment can be damaging at both a personal and professional level, negatively impacting the players' ability to be and perform at their best.

Abusive comments are personal attacks on the identities and characteristics of players that can have detrimental effects on their overall well-being and can also lead them to hide and withdraw from who they are and want to be.

The consequences of being exposed to continuous abuse online for professional football players across the globe are underestimated and oftentimes understated. Footballers can feel discouraged to admit feeling impacted by social media abuse and may choose to act as if all is well with them. These self-protective strategies can place players at even greater risk of not addressing the impact of online abuse on their wellbeing and of not receiving the support they need.

"Social media abuse is a big issue and it's happening all the time. We need to get a handle on it and find ways to stop it because it can have a huge impact on our mental health."

Ali Riley, New Zealand

REAL WORLD SYMPTOMS

The horrific abuse that we see across social media and all online platforms exposes players across the globe to serious risks.

Being victim of online abuse and continuous hatred online can lead to a range of real-world consequences and symptoms, including (but not exhaustively):

- anxiety attacks
- depression
- accumulation of trauma
- low self-worth
- sleep disturbances
- change in eating patterns
- feelings of inadequacy
- social withdrawal and isolation
- and in extreme cases, death by suicide.

"If there is one thing that footballers suffer from the most, apart from losing, it is all the abusive comments – the taunts, the insults. Beyond what we do as professional footballers, we are people. Some players are able to put up with the outrageous abuse we receive online, but other players aren't. It is a very sensitive issue when it comes to mental health."

Leicy Santos, Colombia

FIFPRO

FOOTBALL PLAYERS WORLDWIDE Analysis on the mental health toll on Players contributed by FIFPRO.

NEXT STEPS – LEARNINGS + PREPARING FOR FUTURE EVENTS

FIFA Women's World Cup Australia & New Zealand 2023™ represented the seventh of nine events and tournaments in the first full year of operation for FIFA's SMPS. Valuable learnings and insights have been gathered allowing improvements to the service in terms of tactics, filter words and emojis.

With a broad range of issues and specific insights around the targeting of players and officials, this study offers numerous learnings that have been integrated into the SMPS service, further evolving its reach and effectiveness.

1. Sexist, sexual + homophobic abuse

- Making up almost 50% of all detected abuse across the tournament, Sexist, Sexual and Homophobic messages appear to be the weapon of choice to target players.

- In studying the reported messages provided by this exercise, platforms can strengthen their own filters and moderation libraries to build on existing protections.

2. Player support + outreach

- The data captured in this report displays the effectiveness of proactive monitoring and early warning to trigger outreach to players who are the targets of a pile on (high-volume of targeted abusive messages).

- The ability to provide moderation support to players under these circumstances at the earliest possible moment, (both directly and via their Member Associations), can have a tangible effect on the players performance and mental health.

3. Investigations + action

- Working with Member Associations to agree strengthening of protection via native platform tools, and any action available via the SMPS would enhance future success.

- Sharing learnings – particularly on the issues and phrases used to target players and Member Association official accounts – would better prepare Member Associations to protect their own player's – via Moderation services and switching comment settings off on player channels during event periods.

4. Political activism + support

- There has been a noticeable increase of targeted abuse connected to political statements of support or commiserations.

- Politicians' posts / comments, triggered large spikes of abuse from their own political opponents.

- Using the data from this study, it would be possible to brief political party / leader offices that these activities can have a negative knock-on effect.

5. US + politically motivated abuse

- With the largest segment of politically motivated abuse coming from the US market, it's interesting to note that the next Men's World Cup will take place in the same region, crossing over with campaigning for the US Midterm elections.

- Briefing campaign / US authorities and political operatives of the knock-on effects from senior politicians sending support to the national football team could be a useful preventative measure given the location of the next Men's World Cup in 2026.

6. Match Officials + technology

- Data identified in this study illustrates a concerning rise of abusive content targeting match officials. With a heavier reliance upon technology (including VAR and the introduction of microphones for referees for transparency of decision making), this is likely to be an ongoing issue.

- Briefings for Match Officials ahead of tournaments to recommend increased protective measures (including moderation services) can provide enhanced protection.

THREAT MATRIX

FIFA SOCIAL MEDIA PROTECTION SERVICE
WOMEN'S WORLD CUP 2023™ – ANALYSIS

PROACTIVE MONITORING SCOPE

WHAT WE'RE DOING

FIFA does not tolerate abusive, discriminatory or threatening behaviour targeting officials, players, coaches, staff or their family members.

The SMPS service **moderates** (i.e. to hide or delete, varying by mechanics of the platform in question) abusive, discriminatory and threatening content including but not limited to text, emojis and images.

Where platforms do not support moderation, FIFA also proactively **monitors** social media for abusive, discriminatory and threatening content.

FIFA has and will continue to take action against individuals or groups who are evidenced to have produced or disseminated abusive, discriminatory and threatening social media content, including **reporting** offenders to social media platforms and working with regional law enforcement where appropriate.

Content will be subject to moderation, monitoring or reporting where it includes a reference – whether express or implied – to any one or more of the reasons listed by Article 4 of the FIFA Statutes, where the context may be reasonably concluded to be harmful. Additionally, any content which may be deemed to include threat of harm to the subject or their family members will be automatically included for assessment.

Abusive, discriminatory and threatening behaviour has no place in football.

WHAT WE WON'T DO

FIFA will not use private data or seek to compromise platform-based privacy settings.

FIFA will not deploy surveillance technology to monitor individuals – both moderation and monitoring are issue-based, not individual-based.

FIFA is a strong advocate of free speech and this policy is only designed to tackle abusive, discriminatory and threatening content.

ARTICLE 4 OF THE FIFA STATUTES

"Discrimination of any kind against a country, private person or group of people on account of race, skin colour, ethnic, national or social origin, gender, disability, language, religion, political opinion or any other opinion, wealth, birth or any other status, sexual orientation or any other reason is strictly prohibited".

GLOSSARY

- **Abuse / abusive posts and accounts:** Refers to content that includes verified discriminatory, egregious and aggravated terminology.

- **Discriminatory flags:** Posts flagged for content that may include racist, homophobic, sexist etc. terms.

- **Dogwhistle:** An abusive message clearly exploiting a racist trope without explicitly using directly racist language.

- **Fan identification:** Info on profile that indicates with high likelihood a user supports a particular country / domestic football club.

- **Flagged posts:** Posts flagged for content that may include personally abusive or discriminatory content. Personally abusive content can include calling someone a a**hole, etc.

- **GRT:** Gypsy Roma and Traveller communities.

- **Inclusion criteria:** A post will have mentioned one of the monitored player handles or one of two key terms (or variations thereof). Multiple reports can be generated from a single post / comment if more than one offence is detected.

THREAT MATRIX

FIFA / FIFPRO

SOCIAL MEDIA PROTECTION SERVICE

FIFA WOMEN'S WORLD CUP
AUSTRALIA & NEW ZEALAND 2023™

FULL TOURNAMENT ANALYSIS

Signify | respondology