

KEY FINDINGS: ONLINE ABUSE IN ATHLETICS

AI RESEARCH STUDY:
4-YEAR COMPARATIVE INSIGHT ANALYSIS INTO ONLINE ABUSE OF
ATHLETES DURING GLOBAL ATHLETICS CHAMPIONSHIPS





INTRODUCTION

World Athletics engaged Threat Matrix over a four-year event cycle to study online abuse targeting its athletes at the Tokyo 2020 Summer Olympic Games (held in 2021), the World Athletics Championships in Oregon 2022, World Athletics Championships in Budapest 2023 and the Paris 2024 Olympic Games.

This report brings together key data, insights and learnings from all four events in this first-of-its-kind comparative study in global sport. As the first International Federation to conduct such an in-depth, longitudinal study, World Athletics has initiated other International Federations to change their approach in this space.

Threat Matrix is an initiative by ethical data science company Signify Group supported by sports investigations company Quest Global. It uses machine learning and AI Natural Language Understanding to detect abuse and fixated threat online, helping World Athletics understand the issues and present it with actionable, real-world solutions.

World Athletics' deployment of Threat Matrix has developed and expanded in scope and coverage since the initial Tokyo 2020 Olympic Games study. The findings and analysis of each event have allowed for a more in-depth study built on the learnings of the previous events, taking athlete online protection ever further.

By the Paris 2024 Olympic Games, the study covered 1,917 athletes with at least one active social media account - this is almost 12-times the size of the original sample set activated at the Tokyo Olympics (161).

This report presents and analyses the tactics used to abuse and threaten athletes online, and how these have changed over time from one event to the next. The themes of racism, sexism and sexualised abuse, as well as general abuse and trolling, remain the constant themes throughout the four events, a breakdown of which is set out in this report, alongside actions and initiatives taken to enhance online protection for athletes in sports and disciplines governed by World Athletics.

CONTENTS

INTRODUCTION

Executive summary	2
Contents	3
Methodology	4

EVENT COMPARISON

Event comparison	5
Headline figures	6

ANALYSIS & INSIGHTS

Categorisation analysis	7
Platform breakdown	8
Gender breakdown	9
Language & use of emojis	10

TAKING ACTION

Action statistics	11
-------------------	----

CONCLUSIONS

Conclusions	12
-------------	----

APPENDIX

Appendix A: Detailed methodology	13
Appendix B: Glossary / Classifications	15
Appendix C: About this research	16



METHODOLOGY

Across the four competitions, the Threat Matrix service monitored the social media accounts of 2,438 World Athletics athletes in relation to online abuse.

The athletes selected for coverage were provided by World Athletics, along with their account handles, and Signify conducted a verification process to find any additional accounts. Each of the 2,438 athletes covered had one or more account across four social media platforms (X, Instagram, Facebook and TikTok).

The Threat Matrix service was in operation for the entirety of each event, beginning 24 hours before the event and ending 24 hours after the final event and/or closing ceremony.

The four combined studies captured 1.4 million posts or comments containing at least one account handle of the monitored list (on X), or were detected comments posted in response to content published by one of the monitored athletes (on IG, Facebook and TikTok). This study excluded Direct Messaging or non-targeted abuse conversation on X; where the athletes were mentioned but not tagged. This suggests the issue is even greater than the study highlights.

This figure represents the content that contains terminology on the Threat Matrix lexicon, across 40+ languages, that is deemed as potentially containing abusive or threatening content.

Threat Matrix analysts review all flagged content in a two-stage process to ensure the context of the post is verified as abusive or threatening. For example, distinguishing between “I’ll kill you” and “you killed it”.

1.4 M
total posts and
comments captured
for analysis

2,438
athletes with at
least one active
social media account
covered across the
four events

1,258
total posts and
comments verified as
abusive

1 SOURCE DATA
Using AI-powered threat detection algorithm, Signify scan public posts

2 CLEAN DATA
Removing bots, discriminatory abuse targeting selected athletes is flagged

3 ANALYSE DATA
Flagged posts are analysed by Signify’s team of experts

4 EVIDENCE + ACTION
The most egregious examples of abuse are prepared and submitted for action

EVENT COMPARISON

Monitoring carried out at the Paris 2024 Olympic Games consisted of the largest athlete data set covered as part of the World Athletics study so far, with 1,917 athletes covered (all athletes possess at least one active social media handle). This generated 355,873 posts and comments mentioning these active handles. This compares to the previous events as follows:

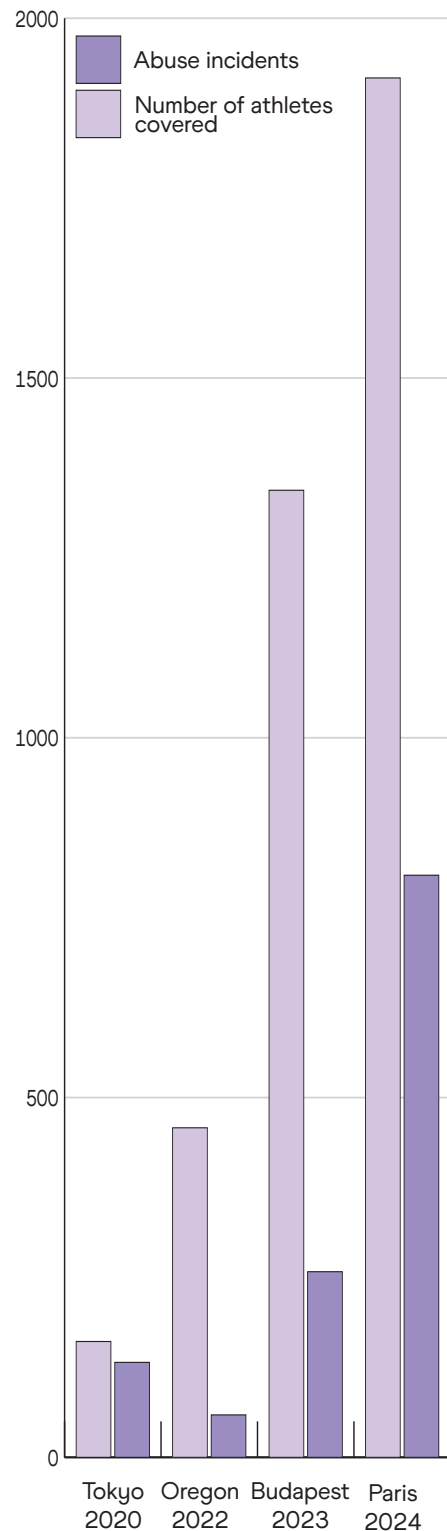
- Tokyo = 161 athletes monitored (240,707 posts / comments*)
- Oregon = 458 athletes monitored (427,764 posts / comments*)
- Budapest = 1,344 athletes monitored (417,240 posts / comments*)
- Paris = 1,917 athletes monitored (355,873 posts / comments*)

Given the different range of athletes and accounts monitored, drawing a detailed comparison between the four events is complicated. However, there are useful trends that can be detected.

Three out of the top five most targeted athletes from the 2023 Budapest Championships also featured among the top five most targeted athletes at the Paris Olympics. This provides a strong indication that outreach to an existing highly targeted athlete could have a positive impact.

The figures for the 2020 Tokyo Olympics were influenced by two athletes receiving almost 65% of all detected abuse, while the Paris 2024 Olympic Games also reflected and even accentuated this, with 84% of all abuse targeting two athletes.

Please note the Tokyo 2020 pilot only covered a smaller pool of higher profile athletes, hence the stark difference in ratio of abuse incidents to number of athletes covered compared to the other events.



* Posts / comments for Oregon and Budapest include monitoring official event handles whereas these were not included for Tokyo and Paris due to high levels of non-athletics related content present on the Olympic handles

HEADLINE FIGURES

TOKYO 2020

Platforms monitored: X

161 athletes covered

240,707 posts captured for analysis

132 posts verified as abusive and reported

63% of abuse was targeted at two athletes (both female)

29% of abuse was sexist

26% was racist abuse

OREGON 2022

Platforms monitored: X @

458 athletes covered

427,764 posts captured for analysis

59 posts verified as abusive and reported

59% of abuse targeted at female athletes

36% of abuse was sexual or sexist in nature

19% was racist abuse

BUDAPEST 2023

Platforms monitored: X @

1,344 athletes covered

417,240 posts captured for analysis

258 posts verified as abusive and reported

35% of abuse was racist – 12x increase compared to 2022 Championships

Unlike previous events, abuse was driven by events outside the stadium and based on athletes' wider reputation

PARIS 2024

Platforms monitored: X @ f d

1,917 athletes covered

355,873 posts captured for analysis

809 posts verified as abusive and reported

18% of all verified abuse was racist

82% of all abuse targeted two athletes

Over **30%** of detected abuse in this study was sexual or sexist in nature

CATEGORISATION ANALYSIS

Across the four-year study there were developments in the Threat Matrix lexicon in terms of language detected and categories of abuse covered.

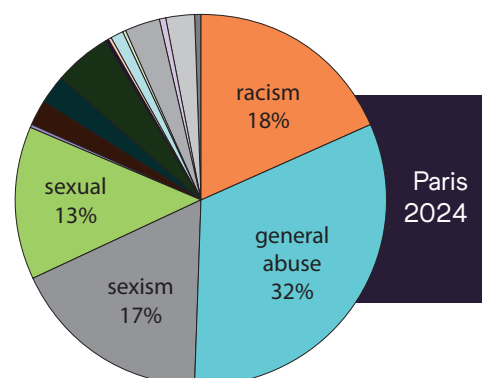
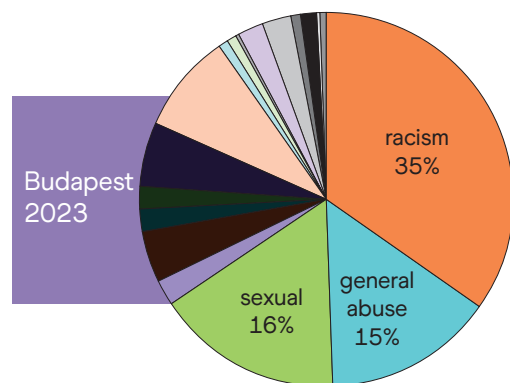
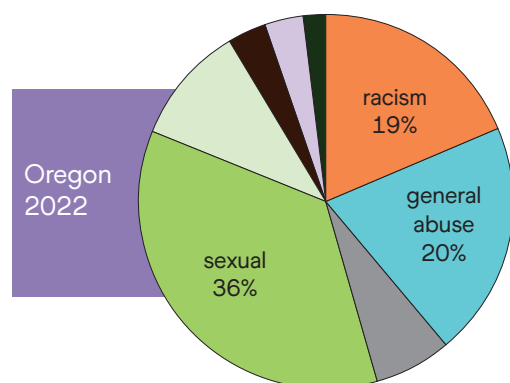
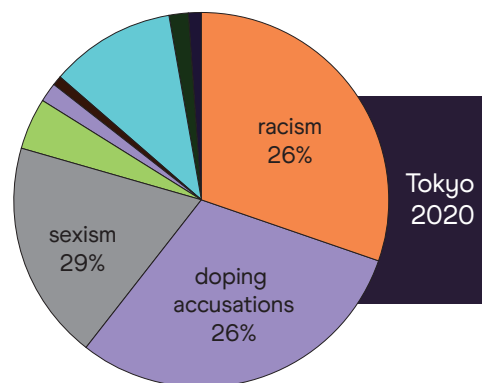
However, throughout all four events racism and sexism consistently remained key issues that athletes were having to deal with online. The final three events all saw general abuse and sexual-related abuse featured as top targeted categories, too.

Racism was in the top two categories of abuse, in terms of detected volume, for all four events. This would make it the stand-out issue and form of abuse used to target athletics athletes.

Sexism featured heavily in all events, being the largest category at Tokyo 2020. It was notable that sexist remarks were sent to both male and female athletes, but comments sent to male athletes of this nature usually mentioned their female family members or partners.

Sexual-related content was also sent to both male and female athletes. However, there was a clear distinction in how sexual terms were used to abuse male and female athletes. For male athletes, sexual terms were used as insults such as “f*** you p****” or “you suck a**”. Whereas for female athletes, sexual-related abuse mentioned more intentional language, such as “I’m going to r*** you”, or references to them being “w*****”.

Detailed abuse categorisation definitions can be found in Appendix B (p.15).



PLATFORM BREAKDOWN

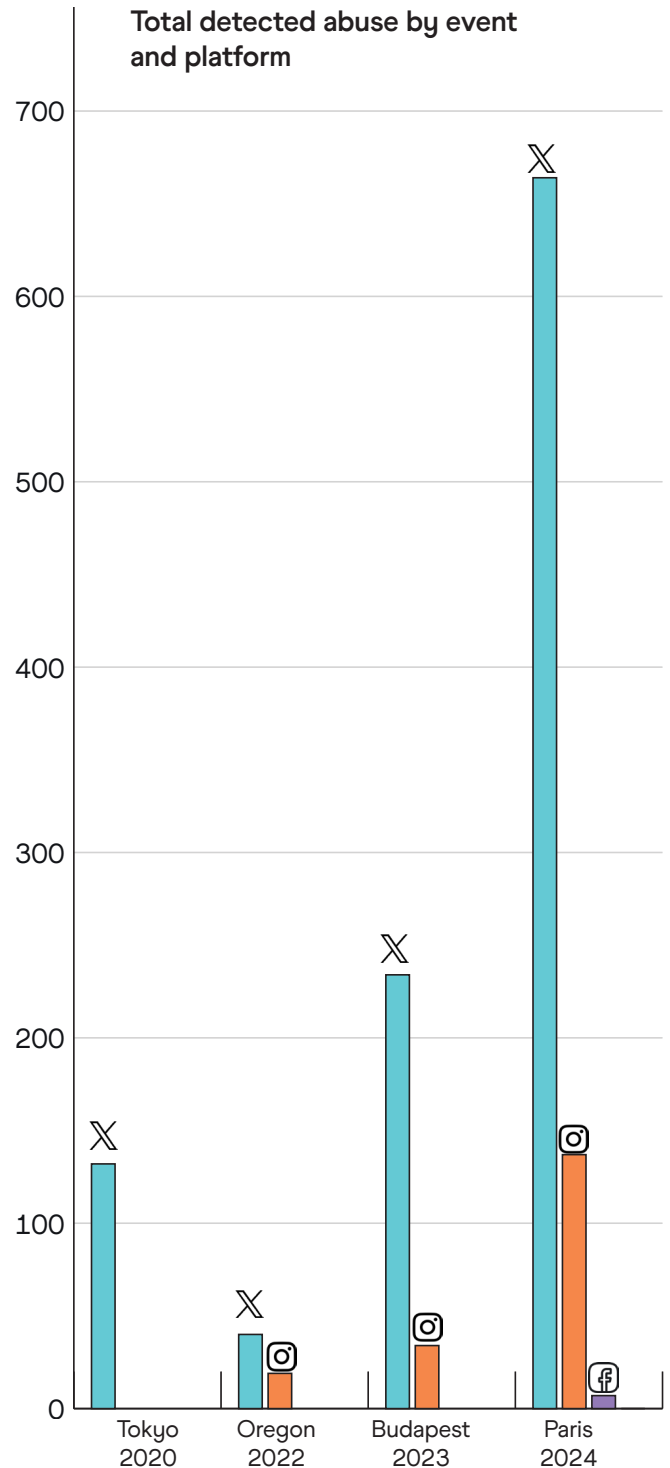
As the scope increased, the platform coverage also advanced to include additional platforms. Instagram from Oregon onwards, and Facebook and TikTok for Paris, although no abuse was detected on TikTok.

A platform analysis found several interesting observations, including:

- X (formerly Twitter) remains the preeminent platform for watching live events and therefore abusing athletes in real time. X however does allow full access to data, unlike other platforms, therefore enabling a more complete and accurate detection of abuse.
- The use of Instagram, Facebook, TikTok and other platforms is growing and evolving among athletes.
- Far fewer athletes post during competition time lending an atmosphere of stasis to their accounts - those that do often only do so to celebrate success - and many employ some form of comment management.

As the study demonstrates, abuse and threats are still getting through to athletes on all platforms, further indicating the need for athletes to be supported in ways the platforms are not offering at present.

Every abusive comment / post detected and verified as part of this exercise (and previous studies) has been reported to the relevant platform on behalf of the athlete.



GENDER BREAKDOWN

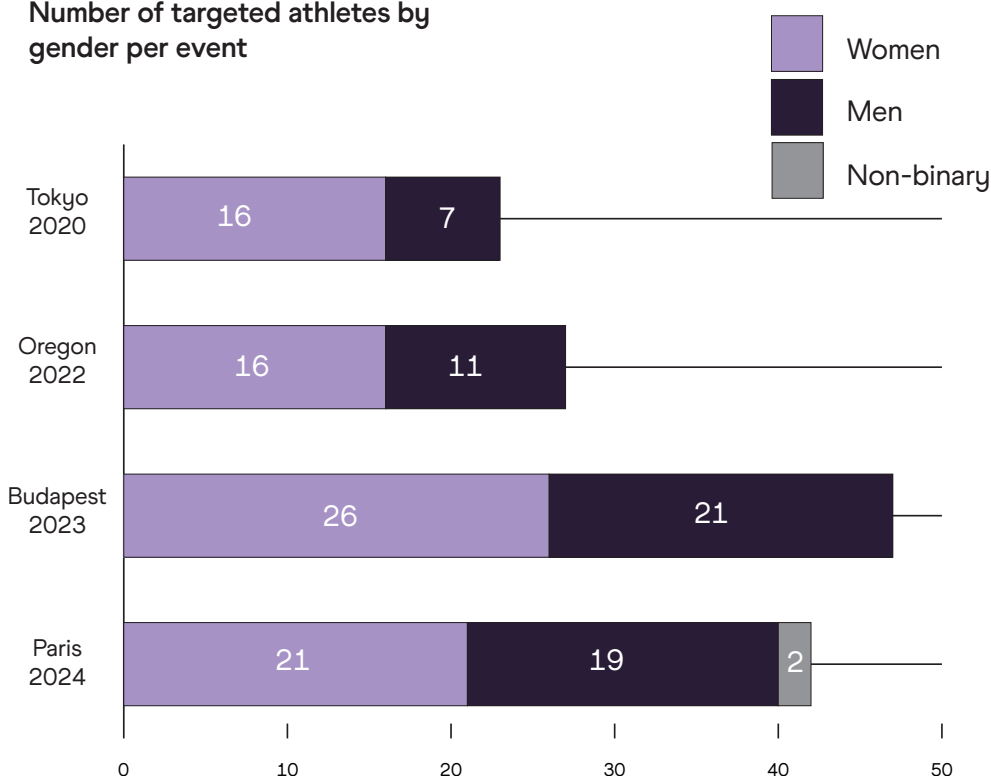
A gender analysis was conducted to assess the differences in volumes of abuse as well as the total targeted athletes.

The analysis highlighted that although female athletes received much greater volumes of abuse at Tokyo 2020, since then the volumes targeting males and females have gradually balanced out.

Volume targeting female athletes at Tokyo was heavily skewed by one female athlete receiving a high volume of abuse from one account, which following investigation, led to legal action being taken. Paris 2024 was the first event where the volume of abuse targeting male athletes was larger – although only marginally – than that which targeted female athletes.

The breakdown of total targeted athletes by gender follows a similar trend to the volume of abuse breakdown previously observed, with initially more female athletes targeted, but a balancing out between male and female athletes targeted across the duration of the four events.

Number of targeted athletes by gender per event



* while competition categories in athletics are based on biological sex, for the purpose of this analysis and to determine the impact of athletes' gender identity on online abuse, non-binary athletes were captured separately in this analysis.

LANGUAGE & USE OF EMOJIS

LANGUAGE ANALYSIS

As the project developed, Threat Matrix was able to introduce a language analysis tool from the Oregon 2022 World Championships onwards.

The results overwhelmingly confirmed that the majority of abuse targeting athletics athletes was sent in English. However, Spanish was also seen in large volumes - particularly during the Paris 2024 Olympics.

The languages of abuse sent appear to correlate with the native language of the athlete receiving the abuse. For example, during Paris 2024 the second most targeted individual was a Spanish speaking athlete, and the majority of Spanish language abuse was targeted toward them.

Overall, English speaking athletes have been the most targeted, hence the high volumes of abuse in the English language (Note: monitored platforms are English-first, meaning that even users whose first language is not English will use that to post. Especially if it is the language of the athlete or official channel).

EMOJI ANALYSIS

Use of emojis in an abusive context was detected throughout the study. Out of the 809* abuse cases, more than 563 emojis were present in posts (with some posts incorporating multiple emojis).

Across X, Instagram and Facebook, we found more than 563 total emojis with a wide range of types. The most widely used emoji categories are animals. These are used for both expressing negative views towards athletes (i.e. rat, snake and pig emojis) or even conveying racist imagery (i.e. monkey and gorilla emojis).

The use of emojis in online abuse has increased significantly, on multiple platforms. This represents a 361% increase in the detection of emojis within verified abusive posts / comments (563 detected at Paris Olympics vs 122 detected at the World Championships in Budapest 2023).

We have seen an exponentially large increase in emoji use since the World Championships in Budapest 2023 and Oregon 2022. This data indicates that the 'weaponisation of emojis' has continued to increase, as emojis are often used to circumvent platform guidelines, being deployed with not immediately obvious context (i.e. corn or aubergine to symbolise pornographic content).

* these numbers exclude data from 2021 as Threat Matrix did not capture data on emojis during the Olympic Games in Tokyo 2020

TAKING ACTION

As well as providing insight into abuse and threats targeted at athletes, teams, coaches, officials and other stakeholders, Threat Matrix is designed to facilitate real world action.

Over the course of the four events, the following action was taken:

- All 1,258 posts were reported to the platform as abusive for action
- 254 accounts directly escalated to platforms for elevated account action
- 2 serious cases have been highlighted to authorities after the accounts had been identified and evidence gathered
- 25 athletes have been offered year-round AI protection on some of their social media platforms after having been identified as highly-targeted individuals during major events – this will be expanded over the next year(s)
- Educational materials have been created and distributed to athletes to provide more information on how to better protect themselves and their accounts online and more are currently in development

The legacy of this action has been to remove a significant number of abusive accounts, preventing them from targeting athletes at future events.

The four-year, multi-event study has given World Athletics and the Threat Matrix team a unique perspective and the data insights necessary to outline a path for World Athletics to further enhance online protection of its athletes. And as important, to heighten the athlete care system to ensure the effects of online abuse are minimised and mitigated.

Using this invaluable data, World Athletics is now working on an action plan that will support its safeguarding policy and is exploring the depth and type of future protection for athletes. In order to ensure the most effective deployment of services, a consultation with athletes on this topic is also taking place.

1,258
posts reported to
respective platforms

254
abusive accounts
directly escalated to
platforms

52
actionable cases with
positive IDs

2
cases pursued under
direction of World
Athletics

CONCLUSION

This ground-breaking four-year research initiative conducted by World Athletics into online abuse during major sporting events represents a significant step toward addressing a pervasive and often underreported issue.

The extensive study has not only shed light on the patterns and prevalence of online abuse faced by athletes during peak moments of their sporting careers, but also highlights the disproportionate impact on individuals based on gender, race and other personal characteristics, as well as the impact of events outside of the sporting arena.

By rigorously analysing data over multiple events, the research has successfully built a clearer understanding of the abuse landscape, allowing for evidence-based discussions and fostering greater awareness among stakeholders.

An essential aspect of this work was to ensure that athletes can confidently and safely engage with social media platforms. World Athletics has invested in and put great emphasis on empowering athletes with knowledge and tools to grow their online presence.

Social media serves as a vital space for athletes to connect with fans, share their journeys and build personal brands that extend beyond their competitive careers. However, the benefits of this engagement must not come at the expense of their mental health or sense of safety. By addressing online abuse, World Athletics reinforces its commitment to enabling athletes to leverage social media as a positive and empowering tool. Ensuring a secure online environment allows athletes to focus on authentic storytelling and professional growth, fostering a healthier and more supportive digital ecosystem for the entire sporting community.

One of the key successes of this initiative has been its role in guiding proactive measures to mitigate online abuse. The insights gained have informed the deployment of tools, policies and collaborations aimed at creating a safer digital environment for athletes.

Additionally, the research underscores the importance of collective responsibility, urging social media platforms, governing bodies, fans and broader society to address online abuse with urgency and accountability.

This ground-breaking work sets a strong foundation for future action, emphasising the need for continued vigilance, reporting mechanisms and robust education campaigns to combat abuse. By leading this effort, World Athletics has positioned itself at the forefront of safeguarding athlete well-being, reinforcing that no one should face hostility for their participation in sport. This initiative serves as a powerful reminder that understanding the problem is the first step toward meaningful change.

APPENDIX A: DETAILED METHODOLOGY

CHANNELS AND MEDIA FORMATS

This study focuses on discriminatory abuse and threats captured on X / Twitter, Instagram, Facebook and TikTok in a range of media formats:

- Text, including word matches denoting abuse or threat
- Emojis
- Images, whether symbolic or text within images
- Voice notes

DEFINITION OF ABUSE

Our definition of abuse is based on the inclusion of a reference, whether express or implied, to any one or more of the following: ethnic origin, colour, race, nationality, religion or belief, gender, gender reassignment, sexual orientation or disability.

As a priority, we look for any threatening comments and those which take a similarly abusive or threatening aspect towards an athlete or an athlete's family.

For this study, we also incorporated additional categorisations of possible abusive terminologies, associated specifically to athletics. These include narratives around doping allegations and transphobia.

SCOPE

The primary focus of this study has been public posts on X (formerly Twitter) and public comments made to selected athletes on Instagram, Facebook and TikTok.

Guided by World Athletics, Signify worked with a selection of 2,438 athletes and officials / media personalities, (all of whom had at least one active social media monitored account). This list of verified accounts / athletes was drawn from a set of athletes provided by World Athletics.



SCALE + COVERAGE

Signify examined 1.4M posts across X / Twitter, Instagram, Facebook and TikTok matching inclusion criteria which differ by platform due to the availability of data:

- Coverage of X (formerly Twitter) includes any posts mentioning an athlete by handle.
- Coverage of Instagram, Facebook and TikTok includes comments posted in direct response to an athlete on any post made for the duration of the event (or covering their recent posts if not published during the event).

Abusive content and posts are detected by our text analysis algorithm, which flags content on the basis of more than 1,000 keywords, phrases and emojis in over 40+ languages.

This flagged 1,258 posts for further review, which were then individually assessed by analysts to see if they met abuse criteria.

In addition to this, we use an AI-empowered threat algorithm to determine posts that contained explicitly threatening language.

LANGUAGE

The Threat Matrix service operates a multi-lingual filter set, with specialist keywords (as well as emojis), representing different categorisations of abuse, discrimination and threat. The service includes 40+ languages (including, but not limited to, the following):

Afrikaans	Danish	Hungarian	Serbian
Arabic	Dutch	Indonesian-Bahasa	Slovakian
Argentine Spanish	English	Italian	Spanish
Australian Slang	Filipino	Japanese	Swedish
Bahasa-Malaysia	French	Korean	Thai
Brazilian Portuguese	French Canadian	Norwegian	Turkish
Chilean Spanish	Georgian	Persian	Vietnamese
Chinese (Simplified)	German	Polish	Xhosa
Chinese (Mandarin)	Haitian Creole	Russian	Zulu
Croatian	Hebrew	Samoan	

APPENDIX B: GLOSSARY - CLASSIFICATIONS

Ableism: Discrimination that devalues and marginalises individuals with disabilities.

Anti-GRT: GRT refers to Gypsy, Roma, Traveller and abuse flagged as such uses language that discriminates against the community.

Antisemitism: An abusive message discriminatory against the Jewish community.

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

Doping: Content containing abusive accusations of drug use to enhance performance. These messages frequently intertwine with transphobic tropes, due to the athletes' alleged use of hormones (testosterone).

Family: Abusive messages targeting or mentioning an athlete's family or partners.

General Abuse: Messages that are abusive in nature but do not contain any specific discriminatory language.

Homophobia: Prejudice, discrimination, or hate towards individuals who identify as gay, lesbian, bisexual, asexual, or questioning. It also encompasses harmful stereotypes related to sexuality which are applied to individuals irrespective of their sexual orientation.

Intracommunity Slurs: Derogatory insults used within a single community and are directed inwardly. The terms often originate from the community itself.

Islamophobia: An abusive message that contains hatred, fear, or prejudice against Muslims.

Match Fixing / Corruption: Abusive allegations of match fixing or corruption, or abusive content posted by angry gamblers.

Political: An abusive message targeting an individual based on their political beliefs or making presumptions on their beliefs as a basis for attacking them. Messages often falsely link an individual to a political issue or event to justify abusing them.

Racism: A discriminatory, derogatory, or hateful message against an individual based on their race or ethnicity, which is typically marginalised.

Sectarianism: An abusive message against an individual who is linked to a religious and political group, typically arising from an ongoing conflict between the specific group and another.

Sexism: Messages that contain discrimination, prejudice, and harmful stereotypes based on sex or gender, typically against women.

Sexual: An abusive message that subjects the targeted individual or group of individuals to graphic sexual language or sexual acts without their consent.

Special Terms: An abusive message that can only be applied to an individual or group of individuals due to specific circumstances.

Terrorism: A message that contains words that are suggestive of, or related to, terrorist activity.

Transphobia: Messages that are hateful, fearful, or discriminatory against non-binary, gender non-conforming, and transgender individuals. In sports, transphobic and sexist tropes often intertwine falsely equating biological sex with gender to make derogatory comments regarding physical strength.

Violence: Messages containing online threats (including death threats) that demonstrate an intent or desire to cause harm to an individual or group of individuals.

Xenophobia: Fear or hatred based on an individual's country of origin.



APPENDIX C: ABOUT THIS RESEARCH

ABOUT WORLD ATHLETICS

World Athletics is the international governing body for athletics (track & field and running events). Our vision is to use the power and accessibility of athletics and our athletes to create a healthier and fitter world. World Athletics serves a number of stakeholders: elite athletes, their coaches and their agents, and the 214 national athletics federations that make up the organisation, commercial partners and broadcasters. We work closely with the national federations to develop the sport around the globe, and with local organising committees to stage the biggest international athletics events, exciting and engaging fans of all ages.

DATA ANALYSIS TEAM

This report was commissioned by World Athletics. The data gathering and analysis was conducted by ethical data science company Signify Group (www.signify.ai). With a specialist capability in the identification of hate speech and social media abuse, Signify have worked with sporting governing bodies and clubs in professional football, rugby and cricket, and have developed a proprietary AI driven monitoring service to protect athletes, clubs,

INVESTIGATIONS TEAM

Quest Global (www.quest.co.uk) run human Intelligence-led investigations using sophisticated, police-led techniques. They can provide law enforcement liaison and bundling / presentation of detailed evidence packages. In this study Quest have supported on the review of abusive materials and initial triaging of accounts.

The Quest team are on standby to run deeper investigations into any of the accounts highlighted in this study, alongside packaging any required evidence.