



# AI Research Study: Online Abuse and Project Restart



Powered by  
Signify 

## Executive Summary

This study has been commissioned by the Professional Footballers' Association Charity (PFA Charity) and supported by Kick It Out. It was carried out by Data Science company Signify Group and looks at targeted, abusive messages sent via social media to 44 high profile current and former players from across the top divisions of English Football (incl. five players currently playing in Germany and Italy for comparison). This study covered the six weeks of 'Project Restart' – the delayed resumption and conclusion of the 2019/2020 season.

The study is a pilot for a wider scheme to encompass more players and more channels. It is intended to demonstrate the validity of using machine learning to capture, analyse and quantify online abuse – and to shine a light on a tiny sample of the hatred directed at footballers during the course of their careers.

Signify's machine learning systems analysed 825,515 incoming messages and ran a deeper analysis of over three thousand explicitly abusive messages. This data provides three key observations. This report explores these, offering evidence and solutions.

**1. Targeted abuse of footballers on social media has become normalised**

**43%** of Premier League players in this study experienced targeted and explicitly racist abuse on public Twitter.

**2. Speaking up can create negative repercussions for players**

**50%** of the total online abuse recorded was received by 3 players who called out racial abuse during Project Restart.

**3. Social media platform blindspots are failing victims of abuse**

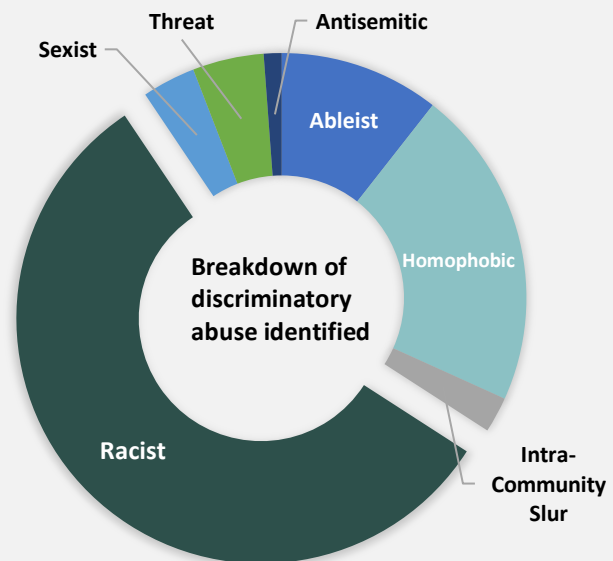
**29%** of racially abusive posts came in emoji form. These posts have not been deleted or accounts banned. This highlights a blindspot for platforms.



## Methodology

The period of study covered the duration of Premier League's Project Restart – 17<sup>th</sup> June until 26<sup>th</sup> July 2020.

In the wake of various pieces of isolated reporting, this project aimed to build a more cohesive picture of the social media environment faced by people in football. One objective of this study was to provide an accurate comparison, looking at public social posts only, and comparing messages directed individually at each of the selected subjects – cutting out spam and crossfire conversations / banter between fans.



The data set for this study (over 825,515 posts across six weeks) is designed to differentiate the signal from the noise surrounding high profile players. The worst, most threatening messages are sent via private, direct channels. However, the use of only public posts in this study allows a meaningful comparison between players, and around various incidents.

This report focuses on public posts on Twitter. While not used by everyone, Twitter provides a useful basis for understanding public social interactions. It is also universally used by players and ex-players to manage their brands online.

After separating out 11,000 potentially offensive or controversial messages, a categorisation process was used to highlight patterns of discrimination. This study then looked in detail at types of messages to compare levels of support for different kinds of behaviour. The study investigated the nature and quantity of various kinds of abuse, and how the public reacted to players who spoke up about issues, or about actual abuse.

## Report team

This report was commissioned by the PFA Charity and supported by Kick It Out. The data gathering and analysis was conducted by ethical data science company Signify Group ([www.signify.ai](http://www.signify.ai)). With a specialist capability in the identification of hate speech and social media abuse, Signify have worked with governing bodies and clubs in professional football, and have a proprietary AI driven monitoring service to protect clubs, players, officials and fans from online abuse [www.threatmatrix.ai](http://www.threatmatrix.ai).



## Finding 1. Targeted abuse of footballers on social media has become normalised

17 of the 44 people studied (39%) received targeted, explicit, racist abuse in one single six-week period. For players in the Premier League this rises to 43% (13 out of 30) players receiving targeted, public, racist abuse during Project Restart. Beyond this, 6 people received homophobic abuse, 5 received ableist abuse, 2 received sexist abuse, 2 received intra-community abuse and 1 received antisemitic abuse. **This equates to 45% of the study.**

Some abuse was concentrated – a single player was the recipient of 33% of homophobic abuse – **but almost half of the sample suffered serious online abuse.** In all, 20 players received 85 messages containing targeted abuse or threats that would constitute aggravated breaches according to Rule E3 of the FA handbook.

The actual number of abusive messages is small compared to conventional messages of support and celebration – those messages that qualify as aggravated abuse constitute around 1 in 2,000 posts targeted at the players and ex-players in this study.

Directly abusive and threatening messages constitute a small part of the hostile dialogue that surrounds players. For every explicit attack there are dozens of instances of people accusing footballers of being over-sensitive or lacking a sense of humour.

The sample set suggests that at a conservative estimate, 40% of players are suffering discriminatory abuse from fans on a routine basis.



*20 out of 44 subjects (45%) were sent individual, abusive messages by members of the public in the space of just six weeks. Any of these messages would constitute a sanctionable offence if sent in the other direction under Rule E3 of the Professional Footballers' Code of Conduct.*

## Finding 2. Speaking up can create negative repercussions for players

The six weeks of Project Restart provided an unusually clear illustration of a perennial problem – that those who speak out against racism or other forms of abuse are subject to attempts to shout them down and subjected to further abuse.



This push back takes many forms ranging from resentment that highly paid athletes should complain about any aspect of their lives, to bad faith protestations that abuse is imagined or humorous, to further slurs, hatred and threats.

Since the restart, leagues and players have taken various steps to express support for the Black Lives Matter movement.

Many of the players in this study have been individually vocal, and television broadcasts will have displayed support when games were being undertaken. These conditions have allowed us to look at the kind of reaction that speaking out actually provokes.



### Finding 3. Social media platform blindspots are failing victims of abuse

Without context it is difficult for some forms of abuse to be picked up by Artificial Intelligence or even human moderators – as demonstrated in the examples in this section. Flagged abuse can sit on the platform for months if it is not immediately clear to a moderator why a post has been flagged.

29% of racist abuse in this study came in the form of emojis indicating traditional racist taunts used in stadiums around the world. Despite coming from accounts with months or years worth of historic abusive posts none of the posts were deleted, and none of the accounts engaging in this had been banned prior to Signify and the PFA Charity reporting to [@TwitterSupport](#). Even after the account was reported [@SARRAILLE1](#) initially only had 5 tweets removed after 10 days, leaving a host of racist and violent posts live.

This represents a weak point in Twitter's efforts to deal with racism, had Signify not detected the account it would still be posting targeted racist abuse – this study found several accounts who repeatedly abuse players using symbols and images in order to get around Twitter's own community moderating tools.





A look at the historic Tweets of some of the most abusive users shows that this has been a deliberate and successful tactic.

To demonstrate this, the PFA Charity and Signify have selected highlighted examples from @SARRAILLE1. The account used emojis at least 25 times to target racist, and other, abuse at a number of players and clubs between 22<sup>nd</sup> July - 23<sup>rd</sup> August 2020.

The abuse captured from this account targets a range of players in the English Premier League and across other European Leagues. For the purposes of this report, examples including more graphic and pornographic images / media depicting genitalia and explicitly sexual acts have been excluded. These examples expose particularly vicious and threatening behaviours.

This highlights the problem faced in detecting abuse without specific contextualisation. Emojis were not designed with the intent of allowing one person to racially abuse another, and presumably in virtually all contexts usage is benign. However, it is clear that when aimed at a footballer of colour these emojis are racist.



The same account targeted Manchester City and Chelsea with a focus on Black players.

Abuse displayed here is both racist and life threatening – with knife emojis and direct death threats frequently deployed.

The analysis also picked up evidence of a connection to gambling with death threats following potential financial losses or attempting to influence outcomes, player form and performance.





## Case studies

This section describes the actions, and experiences of three high profile players during Project Restart. In the course of this study, three players stood out as receiving more than half of the explicit racist abuse directed at participants – Wilfried Zaha, Raheem Sterling and Adebayo Akinfenwa.

A similar pattern was visible in each case – the player spoke out about specific incidents, and within the next 48 hours received a sudden spike in direct attention, completely dwarfing what is normal for each player (for Sterling such a spike occurred on two separate occasions).

On each occasion, these spikes included several incidents of targeted, explicit racist abuse. For Zaha and Sterling, the periods around these incidents represented comfortably the highest rate of mentions during the study period. None of the players appear to have been offered any protection by social media platforms, although this may have taken place behind the scenes – the abuse remained unmoderated online.

### Case Study 1: Wilfried Zaha

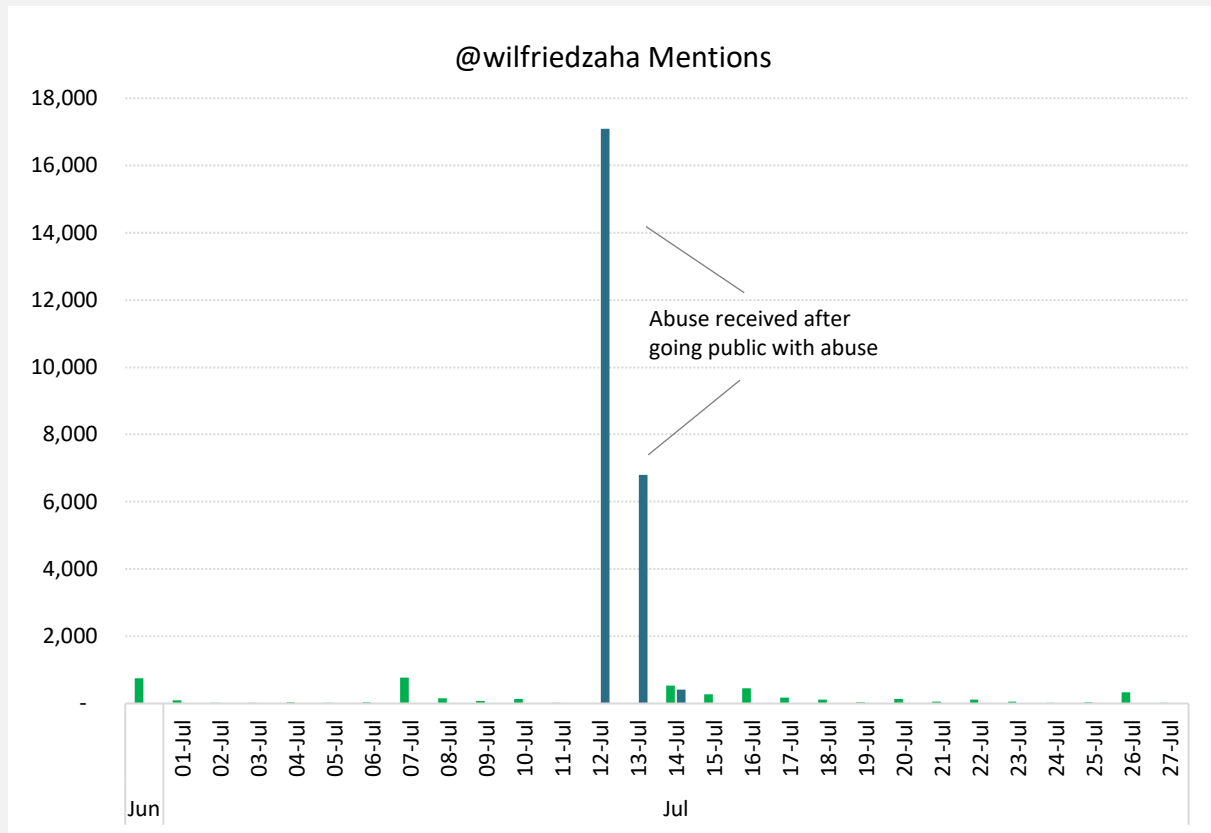
Zaha received several pieces of targeted public racist abuse, all in the period after he reported and publicly complained about abuse in his Direct Messages.

Despite his profile as a prominent professional footballer, only 4,438 posts were directed towards him in the main part of the study period - compared to 24,309 in the period immediately following this tweet.

The abuse he reported was very clear and extreme, with evidence provided in the form of a direct screenshot.

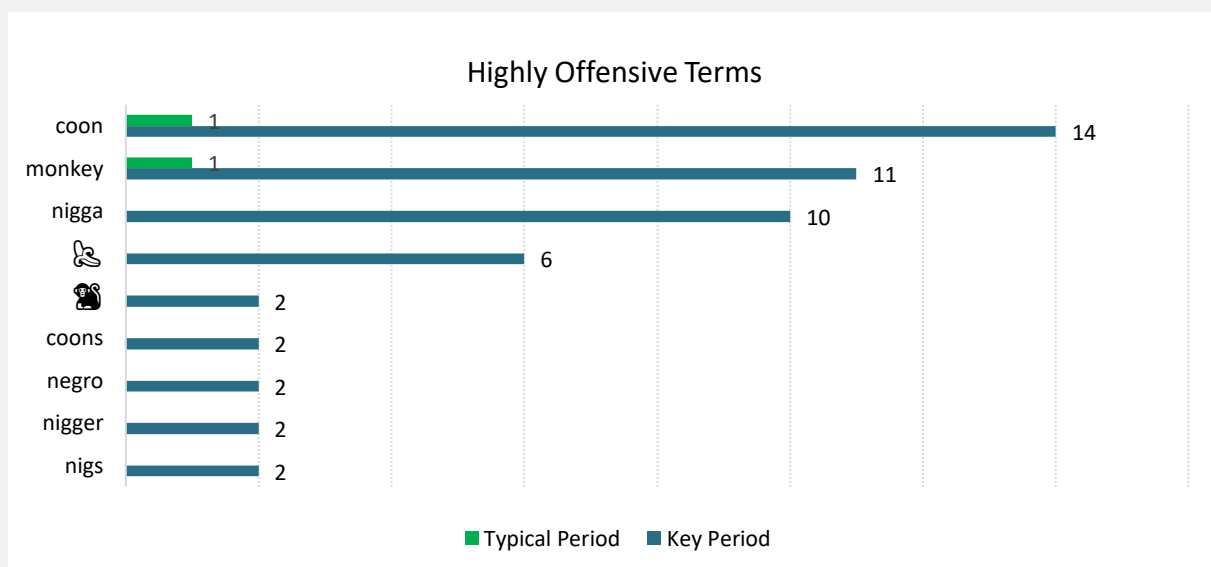
This elicited an overwhelmingly supportive public response – with words like ‘disgusting’ and ‘sorry’ used 2,292 and 1,282 times, respectively.





There were significant counter-narratives, however, which highlight some of the environment in which the use of racist slurs emerges.

The data identified a rise in highly offensive terms, particularly with people repeating what was said in the original message shared by Zaha.





Dozens of posters reduced the experience to a matter of ‘[feelings being hurt](#)’ and accused Zaha of being ‘[fragile](#)’ for being bothered by the abuse.

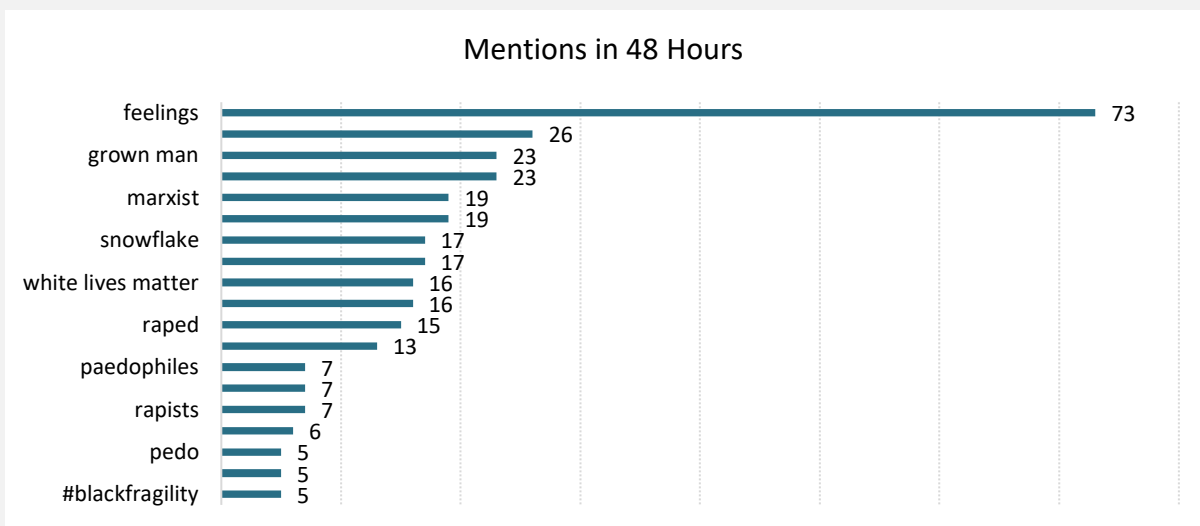
The narrative that Zaha is a ‘[snowflake](#)’<sup>1</sup> fed into the idea that he or others are guilty of ‘[doxxing](#)’<sup>2</sup> a 12-year-old over nothing more than an upsetting message. Zaha is portrayed as a ‘[grown man](#)’<sup>3</sup> attacking a child – a familiar defensive stance for white nationalists.



<https://twitter.com/sperglock/status/1282466185183760384>

In the further conversation ensuing between Twitter, there were 29 mentions of [grooming](#),<sup>4</sup> 30 of [paedophilia](#)<sup>5</sup> and 50 mentions of [rape](#)<sup>6</sup> – a gross subversion of the original complaint by Zaha. Beyond these are the smaller, but familiar tropes where antagonists bring up ‘[anti-white racism](#)’<sup>7</sup> and use it as an opportunity to state that ‘[white lives matter](#)’.<sup>8</sup>

This graph shows the number of times these terms were used in the 48-hour period – whilst some uses of the term are aimed at countering the abuse directed at Zaha, the data evidences how the abuse controls the narrative in this period and acts as a lightning rod for further abuse. The experience of Zaha highlights how people try to ‘talk around’ racism when it appears. In this case, while some outliers make the attempt, the examples shown by Zaha make it difficult for anyone to argue that the content was anything but racist.



<sup>1</sup> [http://twitter.com/1\\_1Shibu/status/1282520252937576449](http://twitter.com/1_1Shibu/status/1282520252937576449)  
<sup>2</sup> <http://twitter.com/RBreslin72/status/1282690309575639043>  
<sup>3</sup> [http://twitter.com/Max\\_Trade1/status/1282394589870137344](http://twitter.com/Max_Trade1/status/1282394589870137344)  
<sup>4</sup> <http://twitter.com/saranti007/status/1282313444893458433>  
<sup>5</sup> <http://twitter.com/Julie98049334/status/1282705263582224384>  
<sup>6</sup> <http://twitter.com/NicholasHowes6/status/1282438594511151105>  
<sup>7</sup> <http://twitter.com/Rwquenu1/status/1282368111568617473>  
<sup>8</sup> <http://twitter.com/steveglasgo/status/1282712685927370752>





Meanwhile many individuals took the opportunity to be directly and explicitly racist in the comments with an air of total unrepentance. This is best understood as part of a wider pattern of behaviour.

The overwhelming majority of people who oppose Zaha make no attempt to deny racism but seek instead to diminish the experience. In the unfolding conversation, there are dozens of people who object to Zaha's decision to report the incident. They seek to turn the tables by portraying Zaha as an aggressor – as someone in a position of power who leveraged this following to harass and persecute a young boy.

The logic is that racist abuse is only experienced as an 'upsetting' comment from a child, clearly ignoring Zaha's claims to have experienced this regularly – and obviously the multiple instances of racist abuse within that very conversation. Not only are there numerous messages to this effect, but some - telling him to '[get a grip](#)'<sup>9</sup> received hundreds of likes.

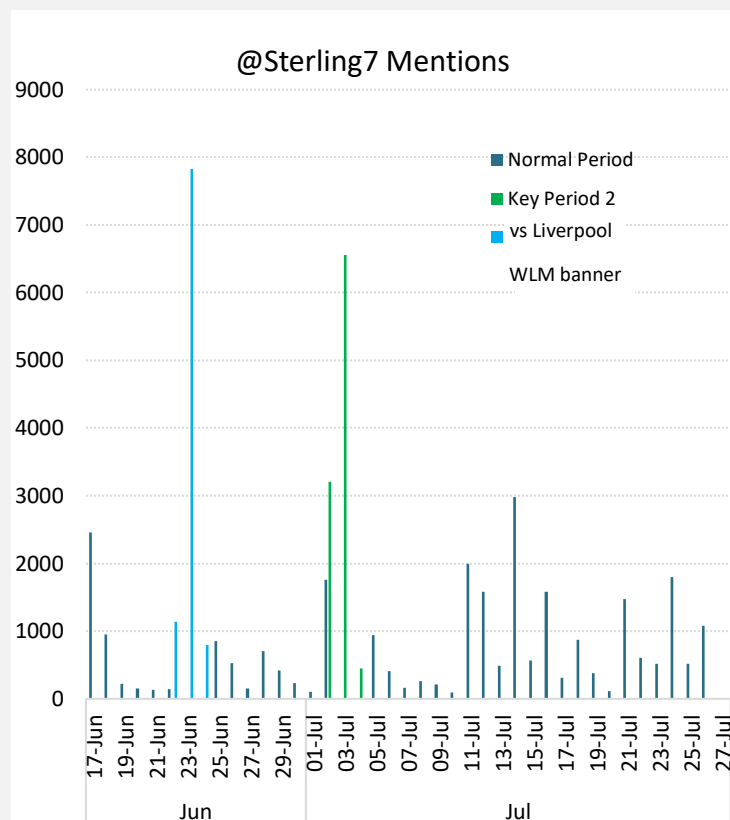
Confronted with clear and obvious racism, many prefer to treat the issue as exceptional or to change the subject.

## Case Study 2. Raheem Sterling

Sterling was the player to receive the largest volume of targeted, explicit and racist abuse of all the players in this study.

During Project Restart, there were 47,737 posts on the platform mentioning his handle.

There were two periods of particularly high interest, and during both of these Sterling received targeted, explicitly racist abuse.



<sup>9</sup> <https://twitter.com/Jim61364753/status/1282359088031178754>



## Manchester City v Liverpool reactions (2<sup>nd</sup> July 2020):

One spike of abuse derived from a post that does not mention race or racism whatsoever.

Here, Sterling talked about abuse following a game against his former club Liverpool on 2<sup>nd</sup> July 2020.

There was a significant rise in his mentions around the game, but this took a huge increase when he posted – somewhat jokingly – about the abuse he was receiving.

In the 48 hours that followed this tweet, Sterling was the subject of 10,215 tweets.



<https://twitter.com/sperglock/status/1282466185183760384>

## Burnley v Manchester City: White Lives Matter banner (22<sup>nd</sup> June 2020)

The more significant incident occurred in a game where Sterling was present but did not feature, as Manchester City beat Burnley 5-0. A banner saying 'White Lives Matter' was flown above the Etihad Stadium, and has been attributed to supporters of Burnley.<sup>10</sup>

Activity spiked when Sterling posted an image of the White Lives Matter banner flown above a game against Burnley.

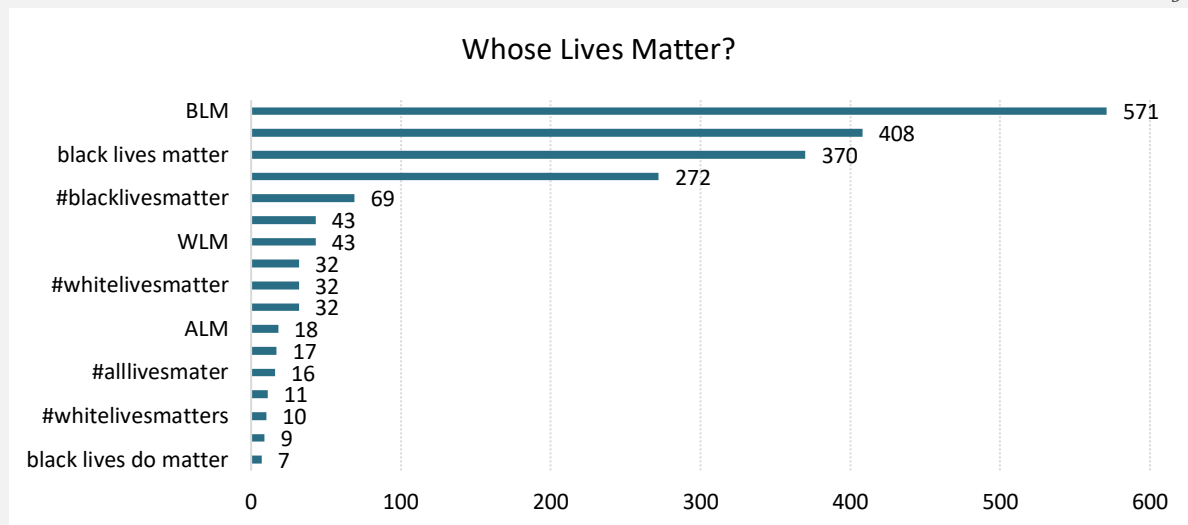
In the 48 hours after this tweet, 9,755 Tweets appeared mentioning his handle.

Predictably, much of the ensuing conversation featured a discussion of the BLM protests and expressions of support or disdain for various statements.



<https://twitter.com/sterling7/status/1275189215630954499>

<sup>10</sup> <https://metro.co.uk/2020/06/24/burnley-fc-fan-behind-white-lives-matter-banned-sacked-employer-12898892/>



‘All Lives Matter’ was the phrase most prominently said as a full statement (with ‘BLM’ often used to refer to protests or organisations).

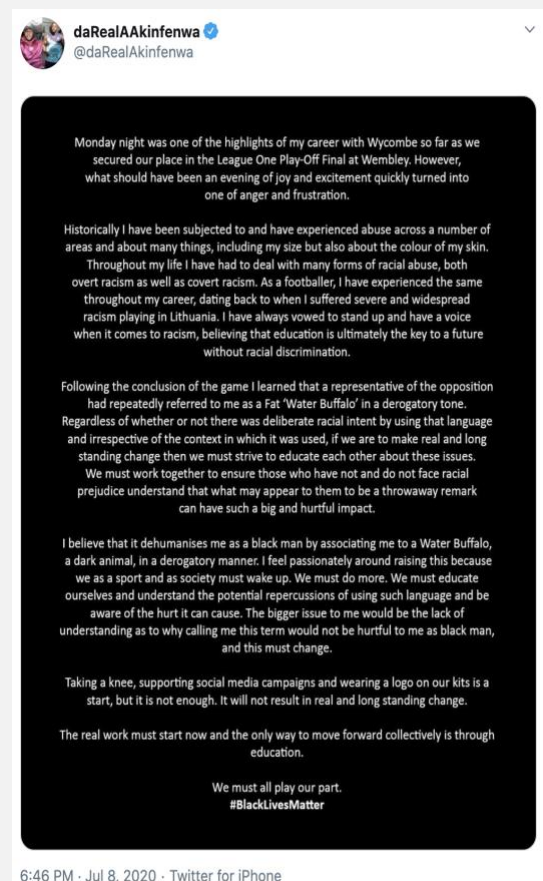
### Case Study 3. Adebayo Akinfenwa

Adebayo Akinfenwa played for Wycombe Wanderers in League One for the duration of the study, gaining promotion via the playoffs on 13<sup>th</sup> July 2020.

Akinfenwa has a significant public profile and will be familiar to a large proportion of fans – this is something that led to him going viral after [Wycombe’s promotion was sealed](#)<sup>11</sup>. He is a fan favourite and gathers significant media attention.

One week prior to the team’s promotion, Akinfenwa experienced a strong backlash after speaking out about racism. On 8<sup>th</sup> July 2020, he posted about an alleged racist incident occurring after a game two days earlier, said to have been perpetrated by a member of the opposing club’s staff.

The focus of this case study is the reaction and patterns identified in response to the players post, as opposed to an assessment of whether the alleged abuse was of a racist nature or not (this was the subject of an FA investigation).<sup>12</sup>



<https://twitter.com/daRealAkinfenwa/status/1280921337893138433>

<sup>11</sup> <https://www.youtube.com/watch?v=geFU0y6f-y0>

<sup>12</sup> <https://www.thepfa.com/news/2020/9/24/pfa-statement-adebayo-akinfenwa>



Following Akinfenwa's post, there was a surge in mentions – going from around 70 posts per day in the previous three weeks to 1,185 in the 48 hours after the report (1,469 posts 17<sup>th</sup> June – 8<sup>th</sup> July vs 1,185 in the following 48 hours). This increase was driven by the backlash in response to highlighting the original abuse.

There were a number of very noticeable shifts in the conversational landscape but most notably, repetition of the original offensive term with more than 10% of his mentions in this period (134) explicitly repeating the term mentioned by Akinfenwa. Despite supportive occurrences, a narrative quickly emerged based on the premise that water buffalo is not a familiar racial epithet.

Another theme revolved around people who [uncovered a joke made by Akinfenwa in February 2020](#) in which he compared the way a teammate jumped on his back to a lion attacking a water buffalo, insisting that this proved the claims vexatious.

Several individuals chimed in to say how they felt it should be taken as a '[compliment](#)'<sup>13</sup> as it makes reference to the player's trademark [strength](#)<sup>14</sup> and [size](#).<sup>15</sup> Other attempts to portray the event as definitively '[not racist](#)'<sup>16</sup> were less subtle.



Direct accusations of [bandwagon jumping](#),<sup>17</sup> acting the [victim](#),<sup>18</sup> playing the [race card](#)<sup>19</sup> and acting in such a way that undermines the seriousness of '[actual racism](#)'<sup>20</sup> abound. In a short time themes like these recurred in over 200 posts – leaving aside people calling him a [fraud](#),<sup>21</sup> an [idiot](#)<sup>22</sup> or an [attention seeker](#)<sup>23</sup> – as racists 'dog-piled' around the incident. The frequency of expletives shows how heated the argument quickly became on both sides.

Posts often contained direct insults of Akinfenwa or other participants, and at least one participant (who appeared to be defending Akinfenwa) chose to engage in [explicit](#)<sup>24</sup> and [repeated](#)<sup>25</sup> [antisemitism](#)<sup>26</sup> towards another commenter. This shows how players outside of the Premier League can also be vulnerable to pile-ons when discussing racism, sometimes more-so as they may lack some of the support available to higher profile players.

<sup>13</sup> <http://twitter.com/MikeD29988538/status/1280946409508941826>

<sup>14</sup> <http://twitter.com/AI3ksandrOrlov/status/1281024882243928069>

<sup>15</sup> <http://twitter.com/MathewForth/status/1281117624680808454>

<sup>16</sup> <http://twitter.com/robmclovintyler/status/1281299989591920640>

<sup>17</sup> <https://twitter.com/TalesGhost/status/1280954030655750145>

<sup>18</sup> <http://twitter.com/yeahyou56/status/1281213274470715392>

<sup>19</sup> <https://twitter.com/lw500efc/status/1280999264014000129>

<sup>20</sup> <https://twitter.com/Richie88571618/status/1281239319827668994>

<sup>21</sup> <https://twitter.com/lufct93/status/1281142128362610693>

<sup>22</sup> <https://twitter.com/Pabsy95/status/1281133822394974208>

<sup>23</sup> <https://twitter.com/pimpjuicedude/status/1280994826788507650>

<sup>24</sup> [https://twitter.com/sergio\\_alferez/status/1281155827714478086](https://twitter.com/sergio_alferez/status/1281155827714478086)

<sup>25</sup> [https://twitter.com/sergio\\_alferez/status/1281159441694547968](https://twitter.com/sergio_alferez/status/1281159441694547968)

<sup>26</sup> [https://twitter.com/sergio\\_alferez/status/1281162882391126018](https://twitter.com/sergio_alferez/status/1281162882391126018)



## Conclusions

Digital and social media plays an increasingly major role in the lives of players and their families, and the development of clubs as global brands. In this world, behaviour of a section of fans (and non-fans) is problematic and appears to be getting worse. Abusers of all stripes are enabled and encouraged by each other, un-sanctioned by social media platforms. Players enjoy little or no protection.

A concerted effort by organisations such as the PFA, Kick It Out, leagues, clubs and social platforms has the potential to have a transformational impact on the culture that surrounds the modern game. The shift in societal awareness around issues in which some abuse against players is rooted offers a unique moment to tackle this problem.

This report shows those who wish to abuse players have been emboldened and continue to believe such behaviour has no consequences. Continued failure to act throughout a moment with such prominent public awareness and debate would serve only to exacerbate abuser's perceptions of invulnerability.

These 'fans' would not behave in this fashion in stadiums, or on the street. It is imperative to ensure that online behaviours carry consistent offline consequences. Players should neither accept nor expect abuse simply for engaging online.

Social platforms and law enforcement are key stakeholders to successfully tackling these issues. It is vital for players, clubs and leagues to get commitment from platforms and law enforcement on engaging and following through with data supplied to them.



## Recommendations

### For Social media companies

- **Recognise and close off blindspots:** Designate Emoji's as a form of discrimination and abuse.
- **Greater use of Artificial Intelligence (AI):** to pro-actively identify images, emojis and other forms of online abuse – building wider reach and smarter contextualisation of abuse.
- **Work with football authorities:** develop an integrated reporting and outcome process to enhance the protection of players - safeguard those who speak out.

### For Government

- **Ensure social media platforms are regulated more effectively:** establish an authoritative and active Ofcom regulatory body with real enforcement powers.
- **A baseline tracker on progress:** use evidence and data to form an ongoing dialogue between government and law enforcement authorities.

### For Football's Stakeholders (Leagues and Associations) and Clubs

- **Co-ordinated and pro-active monitoring of abusive behaviour online:** the adoption of a centralised AI driven tool to proactively monitor abusive users across social media platforms.
- **Unify reporting:** share data across the PFA, Kick It Out, Premier League, English Football League, League Managers Association, The Football Association and Clubs to build a clearer picture of online abuse.
- **Apply offline consequences for online actions:** Aim to identify abusive social media users to apply real-world consequences including prosecution, stadium bans, suspensions within amateur and grassroots football.
- **Recognise and develop a duty of care to the players as employees:** extending beyond the field of play and onto social media.
- **Use data insights to help educate:** providing threat training to players, families, management and staff.

### For Players

- **Report abuse:** use the reporting channels made available to players by the PFA, Premier League and Kick It Out.



## APPENDIX A: List of players in the study

The list was compiled by the PFA and Kick It Out in consultation with Signify's Data Science team. The study aimed to demonstrate that abuse was not defined by club or league by representing all 20 clubs in the English Premier League (EPL) and includes 44 current and former players across the EPL, English Football League (EFL), FA Women's Super League (WSL) and several players playing in Italy and Germany (for comparative benefit). Included in the study are former players, now working in the media and in other parts of the game.

Player	Team	League	Handle
Bellingham	Birmingham City	EFL	@BellinghamJude
Benrahma	Brentford	EFL	@Benrahma2
Taylor	Charlton Athletic	EFL	@lyletaylor90
Akinfenwa	Wycombe Wanderers	EFL	@daRealAkinfenwa
Sancho	Dortmund	International	@Sanchooo10
Lukaku	Inter Milan	International	@romelulukaku9
Koulibaly	Napoli	International	@kkoulibaly26
Balotelli	None	International	@finalmario
Smalling	Roma	International	@ChrisSmalling
Aubameyang	Arsenal	EPL	@Aubameyang7
Mings	Aston Villa	EPL	@OfficialTM_3
Nakamba	Aston Villa	EPL	@Nakamba_11
Solanke	Bournemouth	EPL	@DomSolanke
Jahanbakhsh	Brighton & Hove Albion	EPL	@Alirezajb7
McNeil	Burnley	EPL	@dwight_mcneil99
Abraham	Chelsea	EPL	@tammyabraham
Hudson-Odoi	Chelsea	EPL	@Calteck10
Townsend	Crystal Palace	EPL	@andros_townsend
Zaha	Crystal Palace	EPL	@wilfriedzaha
Richarlison	Everton	EPL	@richarlison97
Morgan	Leicester City	EPL	@Wes5L1nk
Brewster	Liverpool	EPL	@RhianBrewster9
Salah	Liverpool	EPL	@MoSalah
Trent	Liverpool	EPL	@trentaa98
Sterling	Manchester City	EPL	@sterling7
Walker	Manchester City	EPL	@kylewalker2
Greenwood	Manchester United	EPL	@masongreenwood
Lingard	Manchester United	EPL	@JesseLingard
Pogba	Manchester United	EPL	@paulpogba
Rashford	Manchester United	EPL	@MarcusRashford
Saint-Maximin	Newcastle United	EPL	@asaintmaximin
Aarons	Norwich City	EPL	@maxaarons2
Mousset	Sheffield United	EPL	@lysmousset9
Obafemi	Southampton	EPL	@michaelobafemi_
Aurier	Tottenham Hotspur	EPL	@Serge_aurier
Ndombele	Tottenham Hotspur	EPL	@tanguyndombele
Deeney	Watford	EPL	@T_Deeney
Antonio	West Ham United	EPL	@Michailantonio
Traore	Wolves	EPL	@AdamaTrd37
Anita Asante	Aston Villa WFC	WSL	@NicenNeetz
Alex Scott	Former player	n/a	@AlexScott
Eni Aluko	Former player	n/a	@EniAlu
Ian Wright	Former player	n/a	@IanWright0
Sol Campbell	Former player	n/a	@SolManOfficial