FIA / UNIVERSITY

A STRATEGIC RESPONSE TO ONLINE HATE SPEECH IN SPORT

WHITE PAPER

MARCH 2023

FIA.UNIVERSITY

CONTENTS

EXECUTIVE SUMMARY

The Fédération Internationale de l'Automobile (FIA) recognises that online abuse of its athletes, personnel, officials, and volunteers represents a blight on its sport, and has committed to adopting a leadership position in addressing this issue in the motor sport ecosystem, in the first instance, and the wider sporting environment thereafter.

The FIA accepts that to assume such a leadership role, it must avoid mere virtue signalling, instead adopting an approach that will be understood as sustained, committed, and far-reaching, emboldened by a determination to bring about meaningful change through concrete action. In so doing, it welcomes the support of the European Union and many of its member states, alongside many leading figures from other sporting bodies who share its steadfast commitment to enact positive change, aware of the potential future impact of these malevolent activities if left unaddressed.

Consequently, the FIA is no longer prepared to be passive on this issue. The very essence of sport should be that it remains as a free and open setting in which to participate and maintain life-long involvement. The existential threat presented by online hate speech in motor sport, targeted at competitors, FIA personnel and officials, many of whom are volunteers, can no longer be ignored and the Federation recognises that a concerted response is required, collaborating with partners who share our vision of equality of access and who cherish the contribution made by all who value the place of sport in their lives. Specifically, the online threat against our valued FIA female steward, Silvia Bellot, who was the subject of death threats in late 2022, proved a watershed and defining moment for the Federation and, considering this, a decision to adopt a zero-tolerance stance on the matter was accepted.

As a demonstration of its commitment in this regard, the FIA has instigated detailed dialogue with social media platforms, EU and governmental representatives, fellow sporting bodies and other stakeholders operating in this field, to forge effective collaboration and inform joint action. The FIA has committed to mobilising its 244 motoring and sporting organisations in 146 countries across 5 continents in pursuit of this aim, advocating that media, teams, drivers, and fans take a stand against online abuse, too.

Thus, the FIA is of the opinion that coordinated action must be taken now by all previously mentioned stakeholders, operating in a spirit of cooperation, to implement sustainable solutions to this problem grounded within empirically based evidence. To this end, it will invest significant funding to support research via the FIA University, the Federation's corporate education and research facility, to examine digital hate and associated toxic commentary specific to sport. This will provide a much-needed platform for knowledge sharing, education and, in the fullness of time, prevention of this scourge within wider society, let alone sport.

EXECUTIVE SUMMARY

We recognise that we will be the first major governing body of sport to provide a sustained and strategic response of this kind concerning this issue and we call upon all other sporting federations to join our campaign to keep sport social.

As still further evidence of our multifaceted response to the issue of online abuse throughout the sport world, the FIA has partnered with world-leading Artificial Intelligence experts Arwen, as proof of its commitment to offer a strategic response to the blight that this issue exercises in the modern world. Initial work has centred on engagement with those who are the target of these behaviours to ensure anything the FIA does propose has their endorsement, as well as having a realistic appreciation of the limits to which the FIA, independent of other agencies, can achieve on its own. That is why we are pleased to note that other major partners across the motor sport family, including teams competing in Formula 1, have joined our fight against online hate and we look forward to working alongside them on this issue in the time ahead.

Ultimately, the FIA's aim is to ensure sport remains fully accessible and welcoming to all by promoting and safeguarding a respectful environment where everyone can thrive and succeed. Of course, we respect that sport is, and should remain, a cathartic setting for the expression of emotion and a site of intense competition, including on the track. Equally we accept that the interface between the FIA and genuine motor sport fans should be protected, indeed cherished, and enhanced, and in this context, entirely legitimate commentary, including robust, critical forms when justified, is welcomed, as, together, we strive to reach new heights for the sport we love. But there will never be a tolerance level for discrimination, targeted attacks against personnel, volunteer officials and/or competitors, and the FIA will remain resolute in its commitment to tackling this unacceptable aspect of our business across all our activities.

BACKGROUND AND CONTEXT

A systematic review of the relationship between the internet, social media, and online hate speech, undertaken by Castano-Pulgarin, Suarez-Betancur, Vega and Lopez (2021), defines online hate speech as the use of violent, aggressive, or offensive language which is focused on specific sub-groups who share a common identity.

These activities create a power imbalance in which repeated and targeted malevolent commentary has the effect of elevating the vulnerability of its recipients, encouraging their further marginalisation and, ultimately, dissuading them, and those who are like them, from continuing their involvement in their chosen pursuit. Similarly, another useful definition is provided by Kilvington (2021) who frames online hate more specifically as 'spreading, inciting, or promoting hatred, violence, and discrimination against an individual or group based on their protected characteristics, which include "race", ethnicity, religion, gender, sexual orientation, disability, among other social demarcations' (p.258).

This systematic disparagement of a person or group based, typically, upon their ethnicity, 'race', perceived sexual orientation, gender, and/or nationality etc. exercises a real impact on the lives of everyday citizens. Indeed, according to work undertaken by Gagliardone, Gal, Alvez and Martinez (2015) across the European Union, some 80% of people surveyed had encountered some form of online hate, with 40% of respondents claiming that they had been either left frightened or threatened by postings they had read online. The material effect of these experiences is to dissuade genuine commentary from well-meaning, law-abiding citizens.

Consequently, across Europe and indeed worldwide, many sporting bodies and competitors have become increasingly concerned at the growth of online hate speech, with content targeted at volunteer officials, personnel, competitors and, on occasion, fans, proving particularly disconcerting.
Indeed, in September 2022, a statement from Formula 1 teams and drivers condemning online hate speech, reflecting unfounded conspiracy theories relating to racing incidents during Grand Prix events, attracted considerable commentary. As such, in recent years, all governing bodies of sport have been forced to consider and respond to the impact of this malicious activity upon their practice.

Moreover, it is the targeted and unregulated nature of such hate speech, often following high profile and/or contentious incidents in a sporting contest, for example, that gives rise to the most reprehensible forms of communication.

Partly this is a factor of the now ubiquitous nature of modern professional sport, which is widely available across a range of broadcast platforms, ironically including those used, in turn, to mediate online hate against active participants. In association football, female referees Stéphanie Frappart, Yamashita Yoshimi and Salima Mukasanga created history at the FIFA World Cup in 2022, with Frappart leading out an all-female official team during Costa Rica's group stage match with Germany.
However, the toxic commentary that followed this historic moment, mostly of a highly misogynistic nature, only served to confirm that there is still some considerable distance to travel when addressing this issue.

Indeed, Castano-Pulgarin et al. (2021) argue that the main cause of this exponential rise in online abuse over the past three years is related to wider society's contemporary understanding of social deviance. Specifically, the degree of anonymity offered by social media usage ensures a range of activities up to and including serious threats to life can be transmitted without fear of either sanction or retribution on the part of the perpetrator.
It is a situation that if left unchecked could have far-reaching and deleterious effects upon the long-term standing of some sports.

Moreover, the impact of high-profile events, including across wider society, act as 'triggers' for a spike in online hate speech.

Research by Evolvi (2017) into the aftermath of the UK's decision to leave the European Union, commonly referred to as 'Brexit', for instance, revealed a significant increase in the volume of Islamophobic posts, whilst online political discursive patterns in the USA and other recent national settings have involved the proliferation of racist, ethnic and/or gender stereotyping.

SPORT

Whereas historically sports have always provoked emotional responses amongst spectators and competitors alike, there traditionally existed a level of control, aligned with contemporary societal norms, that ensured it remained largely in check. Whilst this meant in some European countries, for example, a concerning rise in fan-related disorder at football matches during the 1970s and '80s, the capacity of authorities to respond to these was more readily apparent. In contrast, the increasing mediatisation of this realm, however, in which sport, entertainment and various forms of media have become intrinsically, even symbiotically, entwined, has forced observers to consider the full extent of this modern relationship and, moreover, the degree to which discourse surrounding sporting events is shaped, even informed, by social media reactions to it. An often-posed question in this regard concerns the motivation of individuals to partake in online hate speech. For many others, because it constitutes such a harmful activity, it is challenging to comprehend the rationale of those who continue to engage in it. Faulkner and Bliuc's (2016) work seeks to posit an explanation for this, claiming that proponents of online hate deploy moral disengagement strategies that somehow allow them to rationalise these actions as being, in fact, those of an alter-ego and thus a vicarious undertaking detached from reality.

Whatever the exact motivations, the increased proliferation of online hate speech is concerning. In respect of this, Willard's (2007) typology of this realm serves to classify the identifiable forms this activity now takes. These include so-called flaming (sending threatening or rude messages), harassment (sending offensive messages repeatedly), denigration (the posting of innuendo or other forms of misinformation), cyber stalking (harassment that include threats to harm), impersonation, outing (revealing information about someone they would prefer to keep private) and exclusion.

Importantly, however, this online hate not only serves to denigrate named individuals and those they are thought to represent, but it can also provoke hostility towards them.

On occasions, the latter may have tragic consequences, with acts of material damage, physical harm, and even loss of life recorded in recent years.

Relatedly, according to Siegel (2020), "Individuals who are close to an online community, or spend more time in communities where hate speech is common, are more inclined to produce hate material" (p.64). As such, it is argued that online hate has become pernicious precisely because it exists outside socially established norms of acceptable behaviour, cultural taboos, or any other concern on the part of the perpetrator of being censored by others. Rather it operates amid a largely unregulated and anonymous sphere, where individuals act without fear of sanction or even identification, espousing views that, under most other circumstances, would lead to their arrest and charge. This willingness to act with apparent impunity, nevertheless, has a caustic and harmful impact on the individuals concerned, the standing of the sport in question, and society at large.

For all this, published research in the field of online hate remains comparatively minimal, even if recent years have witnessed a marked uplift in its dissemination and import. In so far as this reflects the extent of the issue within the public consciousness, this would imply that concern around online hate speech by sporting bodies and the public at large remains a relatively recent phenomenon, even if it is no less impactful because of this.

In the last decade, the rise of open social media platforms, principally Twitter, has also been intrinsic to the growth of online hate speech. Whereas the use of message or chat boards i.e., dedicated sites that permitted users to express views discretely, had historically been the main setting for the expression of both legitimate and malevolent forms of communications, Twitter allowed users who were largely unknown to each other, living in different parts of the world, to exchange opinions in real time. Over time, the use of this facility to comment upon sporting events created both a symbiotic yet potentially troubling interdependency, especially as discourse on 'live' sporting events, overwritten by emotion, often meant such commentary became irrational, hyperbolised and, increasingly, harmful. Whilst legitimate feedback, including criticism, should be accepted by competitors who freely engage in their sport and often financially benefit from doing so, harmful, discriminatory commentary, including threats to physically harm others or inciting acts of violence, correctly reside in an entirely different category.

An important consideration when examining online hate speech is the preponderance of research undertaken through the medium of English and, with this very often, the dominant Western-cultural focus of such published work. As such, studies undertaken in other parts of the world, where English is not a 'first' language, remain isolated and thus represent an important focus for global sporting bodies like the FIA. It is a point recognised by Waqas et al (2019) amongst others, stating that "Almost all the influential studies have been conducted in the context of high-income countries. Research is needed in low and middle-income countries to justify the generalisability of OHR (Online Hate Research) findings as well as to produce culturally applicable interpretations" (p. 17).

Predictably, association football (soccer) is the locus of most work undertaken by scholars researching online hate, followed by American Football. Indeed, the former is not surprising as it is both a global sport and one with an unfortunate history of ritualistic yet all too real fan violence, as well as being a prominent site for the expression of discrimination against minorities and similar forms of deviance. Moreover, there has been some research carried out confirming the extent to which a culture of intolerance that was once only apparent within the context of a football stadium, for example, has now been transposed online and where this has become embedded, it retains the potential to distort these fora as otherwise legitimate sites for critical discourse amongst well-meaning sports fans.

Whilst most incidents of online hate speech are aimed at individuals or groups based on their ethnicity or national identity, commentary designed to incite hatred based on gender and sexual orientation have increased this decade (Dragiewicz et al., 2018). Confirming this, work by Kearns et al (2022) again states that racism remains the most prominent expression of online hate, accounting for almost half of the published academic work currently in the field. Accompanying this was other forms of discrimination, principally misogyny, and, in all cases, it was athletes who were the principal target of this abuse. Their participation in high-profile sporting events often led to an increased form of online hate speech, with the main motivation for this being to delegitimise their involvement and achievements and, ultimately, to discourage others with a similar heritage from emulating their participation. Interestingly, the proponents of hate speech typically self-identify as fans of the sport in question but, despite this, appear to reserve the right to be abusive towards players and officials, including those representing their favoured team/competitor, by often claiming such discourse forms part of mere sport-related 'banter'.

The question of anonymity appears to be central to the ability of hate speech to be sustained online in the face of its widespread revulsion. Despite claims by social media companies that, in most cases, it was possible to identify the perpetrators of online hate, there remains a perception that the absence of robust investigation and prosecution of the individuals behind this remains problematic. The creation of 'in-groups' in which people who hold equally distasteful opinions congregate online and post views that are shared amongst like-minded people offers protection for the perpetuation of this activity, as does a context in which the onus is placed on those offended by such content to report it and, ultimately, pursue restitution. However, others have posited the view that transparency alone will not provide for a remedy to such hateful content whilst still more have argued that it should not be required at all in a setting in which the expression of uninhibited opinion should be protected and, in the view of some, cherished.

SOLUTIONS

Solutions to address online hate content are varied. Expectations are aligned to defined and identifiable groups, such as national governments, governing bodies of sport and, increasingly, social media companies, and largely depend on how the problem is viewed and defined. Very often the recommendations to address this activity include social media education and training, proactive work on the part of sporting authorities to engage with fan groups, workshops with school-aged children, and as part of a broader range of activities designed to protect the welfare of athletes.

The exact outcome of these interventions has rarely been subject to any form of rigorous and detached assessment and therefore how impactful they are remains unclear. It is questionable, for instance, that otherwise well-intended interventions with young sports people will, in turn, reach the individuals posting most of this harmful content and thus the challenge remains one of engaging such people and highlighting the harmful impact of their activities on others, many of whom, it is worth emphasising again, are volunteers.

For a global sporting body like the FIA, understanding the characteristics of online hate speech from a non-Western perspective, thus, remains a priority. It can also reveal forms of discrimination hitherto not as well understood by Western analyses, even if its effect is no less impactful. Alongside the continued focus on racism, the study of other forms of discriminatory hate speech, including gender-based, ethnic, and sectarian, is also necessary amidst an overall more considered and nuanced approach to the topic than has been evident, in the main, to date.

Finally, the significant changes shaping the social media industry itself may also impact its role as a medium for the transmission of online hate speech in the future. The recent (2022) changes in the ownership and organisational structure of Twitter Inc. and the exponential rise in TikTok, especially amongst a younger demographic, will be monitored as being potentially significant in this field. Allied to a fuller understanding of the perpetrators of online hate as unveiled using primary data, there is clearly much more explication of this field to come, including understanding the motives of those engaged in online hate, and the direction of travel for research in this domain. In this regard, the FIA is determined to perform a leading role in advancing work across all these important areas.

PRELIMINARY RESULTS FROM ARWEN.AI

FIA RESEARCH COLLABORATION

Evidence of the FIA's commitment to addressing online hate speech in motor sport is reflected in its decision in September 2022 to collaborate with one of the world leading artificial intelligence (AI) companies, Arwen – one part of a broader strategic response to an ever-growing problem.

Arwen shared the FIA's concern about how social media channels had become progressively more toxic – with volumes of hate rising by 40% since 2019, and volumes of spam bots rising by 350%+ per year since 2013. The impact of this unwanted content goes beyond the negative emotional impact placed upon those who are victims of it, or who witness it. Evidence confirms that 38% of people disengage online when they feel unsafe. This represents more than a third of regular users being lost to toxicity.
As such, on several different levels, it makes good strategic sense to address this issue at the earliest possible opportunity.

The aim of the FIA-Arwen collaboration is to help build safe and inclusive communities, where everyone can contribute in their own way to motor sport free from the fear of intimidation. The drivers who race, the engineers who support them, the personnel and officials who oversee them, and the fans who follow them, all deserve to be able to participate in motor sports online communities free from fear of a toxic online backlash.

In conducting its work, Arwen deploys sophisticated AI and automation to detect and remove toxic social media comments in under a second. After only 5 months of its collaboration with Arwen (September 2022 to January 2023), the number of toxic comments being posted on the FIA's social media profiles has reduced by 66.6%. This indicates toxic posting is gradually being de-normalised – with community members feeling less and less comfortable posting toxic comments. This is what making social media social means and, working with Arwen, the FIA is committed to staying the course on its mission of never being willing to accept discrimination and intolerance concerning its activities, either in person or online.

As part of his commitment to eradicating online hate speech in motor sport, the FIA President, Dr Mohammed Ben Sulayem, provided a lead to addressing this issue by permitting his personal social media channels to be part of the initial pilot research undertaken by Arwen.
Since the company began its work in September 2022, average toxicity on the FIA President's social media accounts has reduced from 15.34% to 10.72%. A reduction of 30.12%. This would indicate that this intervention is, again, de-normalising toxicity on the channel and changing posting behaviour for the better. Arwen is confident that after a further three months the reduction of online toxicity in this instance will

rise still further to a predicted 70%. In the period (September 2022 to January 2023) Arwen autohid 379 severely toxic messages (1.29% of total number processed) confirming that whilst this issue is ever present, its absence from public view will, in time, ensure its intended effect will dissipate.

Over the same period the official FIA social media channels have recorded a reduction in average toxicity, down from 16.56% to 10.16%, a decline of some 38.65%.
Over this period Arwen autohid 3,162 severely toxic messages (0.84% of total number processed), indicating that there is still some way to travel in addressing this issue and, once again, de-normalising online hate speech.

Finally, across all social media profiles (the FIA and President Ben Sulayem) whereas 92.73% of comments were considered safe in September 2022, within five months some 97.57% fell into this category, which represented a 66.6% reduction in "non-safe comments" and is consistent with the FIA's stated ambitions in this regard.

THE FIA'S STRATEGIC APPROACH

The FIA has developed a detailed six-point plan to address online hate in motor sport. It is briefly outlined below and will form the basis of its sustained commitment to tackling this issue in the time ahead.

1. Firstly, to fully appreciate the impact of online hate speech on those harmed by it and, at the same time, implement concrete actions to provide a robust response to it, the FIA will consult with a wide range of relevant individuals, institutions, and agencies, collaborating with world leading research centres to offer an informed, evidence-based approach to the issue.

2. Secondly, the FIA recognises it must work alongside other sporting bodies, representatives of professional athletes and drivers, national governments, and other policymakers, and, importantly, social media companies, to ensure that any proposed actions have meaningful impact, and it will continue its leadership role in this respect.

3. Specifically, the FIA will become the first governing body of sport to launch its own, dedicated research centre into online hate and will appoint leading researchers, Post-Doctoral researchers and provide scholarships to support the work of this centre, which will partner with the FIA University and other global institutions in providing a setting in which peer-reviewed academic publications, White Papers, policy statements, global conferences and other forms of public dissemination will take place.

4. The FIA will be relentless in its campaign to highlight the scourge of online hate on its valued personnel, officials and volunteers and will activate all its communication channels to implement and amplify this approach, bringing forward dedicated campaigns, working with media partners to ensure this message is consistently communicated and, generally, acting proactively to challenge those who, despite all of this, choose to persist with this activity.

5. The Federation recognises that the success of this strategic undertaking will be understood when meaningful change is delivered. We are already seeing the success of an approach that de-normalises toxic hate speech on FIA channels, communicating a message that the Federation will not tolerate this activity and is committed to long-term, strategic investment in supporting this work until this challenge is overcome.

6. We will validate our successes by ensuring this matter remains at the forefront of the FIA's public face, offering consistent endorsement of our gains and availing of every opportunity to highlight our work in this sphere. We will work with the European Union and National Governments and indeed legislatures around the world to lobby and advocate for others to join our campaign to keep sport social and inclusive.

REFERENCES

1. Castaño-Pulgarín, S., Suárez-Betancur, N., Vega, L., and Herrera López, H. (2021) Internet, social media, and online hate speech. Systematic review, Aggression and Violent Behavior, 58, 101608, ISSN 1359-1789.

2. Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N., Woodlock, D. & Harris, B. (2018) Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms, Feminist Media Studies, 18:4, 609-625, DOI: 10.1080/14680777.2018.1447341.

3. Evolvi, G. (2017), "#Islamexit: Inter-Group Antagonism on Twitter", Information, Communication & Society, 22(3) 1–16. https://doi.org/10.1080/1369118X.2017.1388427.

4. Faulkner, N., & Bliuc, A.-M. (2016). 'It's okay to be racist': Moral disengagement in online discussions of racist incidents in Australia. Ethnic and Racial Studies, 39(14), 2545–2563.

5. Gagliardone, I., Gal, D., Alvez, T., and Martinez, G. (2015) Countering online hate speech. Educational, Scientific and Cultural Organization.

6. Kearns, C., Sinclair, G., Black, J., Doidge, M., Fletcher, T., Kilvington, D., Liston, K., Lynn, T., & Rosati, P. (2022). A Scoping Review of Research on Online Hate and Sport. Communication & Sport,. https://doi.org/10.1177/21674795221132728.

7. Kilvington D. (2021). The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate. Media, Culture and Society, 43(2), 256–272.

8. Siegel, A.A. (2020) Online hate speech. Social media and democracy: The state of the field, prospects for reform, pp.56-88.

9. Waqas, A., Salminen, J., Jung, S-G, Almerekhi, H., Jansen, BJ (2019) Mapping Online hate: A scientometric analysis on research trends and hotspots in research on online hate. PLoS ONE 14(9): e)222194.

10. Willard N. (2007) Cybersafe kids, cyber-savvy kids: Helping young people learn to use the internet safely and responsibly. Jossey-Bass.

FIA / UNIVERSITY

FIA.UNIVERSITY