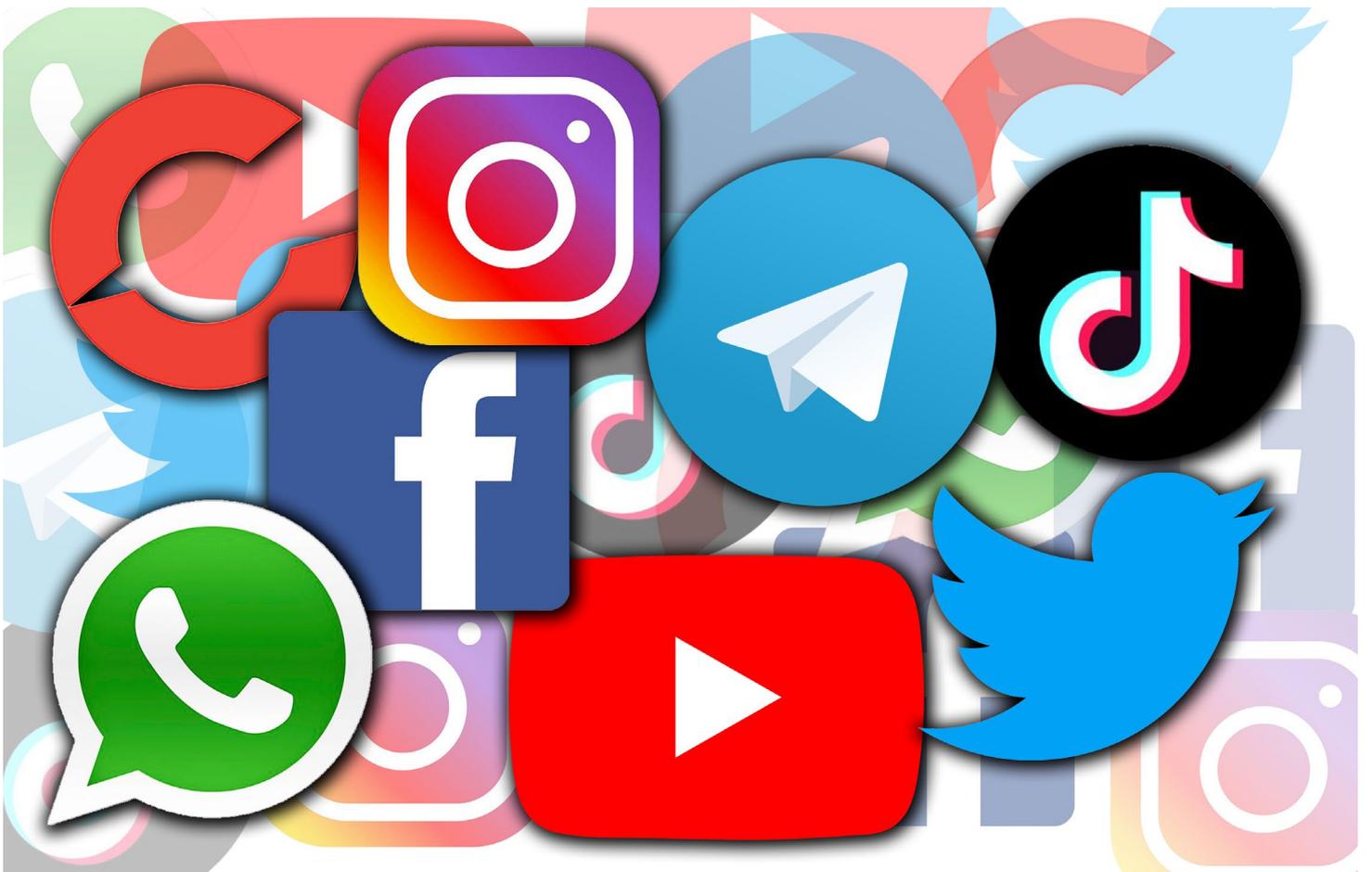


# ONLINE SAFETY BILL

**WRITTEN EVIDENCE SUBMITTED BY HOPE NOT  
HATE TO THE PRE-LEGISLATIVE SCRUTINY  
COMMITTEE FOR THE ONLINE SAFETY BILL**



# CONTENTS

The Time For Regulation Is Now	3
Legal But Harmful Must Continue To Be Included in this Legislation	5
What Is Classed As “Democratically Important” Content?	7
What Is Classed As “Journalistic Content”?	8
Will This Bill Adequately Deal With Small But Extreme Platforms?	9
In Defence of Anonymity	10

# THE TIME FOR REGULATION IS NOW

For years, civil society organisations including HOPE not hate have done vital work on pressuring tech companies to take greater action to reduce harm online. However, it has long been clear that these dangers also require a regulatory solution.

At HOPE not hate we monitor far-right extremists both offline and online and track individuals and groups propagating harms across dozens of platforms all over the world. Different platforms have prioritised dealing with harm to different extents, but it is true to say that almost all have failed to adequately deal with the issue and none have fixed it. The time for legislation is long overdue and so we welcome the Online Safety Bill, which, while not perfect, is an extremely positive development.

As an organisation that attempts to understand and respond to the extremist political landscape in the UK, we are well aware of the importance of online activity to extremists today. Though we campaign against all manner of extremisms, HOPE not hate's expertise and focus lies in tackling the organised far right. As such, our response and recommendations are particularly attuned to how this legislation should undermine the online harms propagated by these actors.

With this in mind, we have limited our response here to just a few elements of the legislation that we feel we have specific expertise on or particular concerns about.

At the same time, we recognise that far-right extremism does not exist in a vacuum and instead emerges from (and feeds back into) wider societal prejudices and inequalities. The activists and groups we campaign against target specific cohorts who are on the receiving end of these systemic prejudices and inequalities, especially women, members of ethnic minority groups, religious minority groups, and LGBTQ+ communities. To the extent that they can be, these wider systemic issues must be addressed in this legislation alongside efforts to curb extremism.

It must also be recognised from the outset that the division between the online and offline worlds is often a false one. All too often what happens online has a huge impact on the personal lives of individuals, but also on our streets and within our communities.

The most obvious example of this is the wave of far-right terrorist attacks around the world,

many of which were birthed online. Many of these terror attacks were carried out by individuals not associated with traditional far-right political parties but rather part of looser, often transnational far-right movements that lack formal structure and emerge online. For most of the post-war period, 'getting active' required finding a party, joining, canvassing, knocking on doors, handing out leaflets and attending meetings. Now, from the comfort and safety of their own homes, far-right activists can engage in politics by watching YouTube videos, visiting far right websites, networking on forums, speaking on voice chat services like Discord and trying to convert 'normies' on mainstream social media platforms like Twitter and Facebook. The fact that this can all be done anonymously hugely lowers the social cost of activism.

This ability to communicate and cooperate across borders would have been inconceivable just a generation ago and while these opportunities are by no means distributed evenly, they have opened up previously impossible chances for progress and development. Yet greater interconnectivity has also produced new challenges. The tools at our disposal to build a better, fairer, more united and collaborative world are also in the hands of those who are using them to sow division and hatred around the world. Legislation, if carried out correctly, can curb the abuse of tools that can provide huge social benefits.

However, this legislation is not just an opportunity to reduce the negative impacts of hostile and prejudiced online behaviour but also a chance to engage in a society-wide discussion about the sort of internet we do want. It is not enough to merely find ways to ban or suppress negative behaviour; we have to find a way to encourage and support positive online cultures. The companies in the scope of the online safety legislation occupy central roles in the public sphere today, providing key forums through which public debate occurs. It is vital that they ensure that the health of discussions is not undermined by those who spread hate and division.

At present, online speech that causes division and harm is often defended on the basis that to remove it would undermine free speech. In reality, allowing the amplification of such speech only erodes the quality of public debate, and causes harm to the groups that such speech targets. This

defence, in theory and in practice, minimises free speech overall. This regulation should instead aim to maximise freedom of speech online for more people, especially those currently marginalised and attacked such as women, Black people, disabled people and LGBT+ people.

Many people underestimate the potential for social inequalities to be reflected in public debate, and disregard the nature and extent of these inequalities in the 'marketplace of ideas'. As such, the position of some opponents of this legislation can be paradoxical. They claim to be committed to valuing free speech above all else, while propagating an unequal debate that further undermines the free speech of those who are already harmed by social inequalities.

# LEGAL BUT HARMFUL MUST CONTINUE TO BE INCLUDED IN THIS LEGISLATION

As an organisation that spends much of its time monitoring hatred online, we are well aware that a huge amount of the harmful content produced by far-right extremists does not reach the legal threshold for prosecution. This content, while legal, can still cause a huge amount of harm. It is for this reason that we strongly welcome the fact that the draft legislation continues to include so-called ‘legal but harmful’ content.

Clause 46 of the Draft Bill defines “content that is harmful to adults”:

Content is within this subsection if the provider of the service has reasonable grounds to believe that the nature of the content is such that there is a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities.

Some regard the inclusion of legal content in this Bill as an untenable threat to freedom of expression. Regulation of our online lives certainly raises legitimate concerns over privacy and freedom of expression. However, as with the offline world, there is a balance of rights to be struck. Government must tackle online harms without infringing on people’s freedoms. It must also preserve freedom from abuse and harassment, and protect vulnerable groups who are targets of online hate. We must end up with a form of regulation of platforms that places democratic rights front-and-centre.

It is HOPE not hate’s belief that regulation must demand changes to platform design and processes that disincentivise legal harm. Illegal content should be treated as a policing problem, dealt with forcefully in law and in court and with the use of appropriate detection and enforcement technologies. Where content violates a platform’s terms of service, it must be removed, and the current enforcement gap must be filled. For a platform like Facebook for example, this would include Holocaust denial, vaccine conspiracism and gendered hate.

However, legal but harmful content, properly defined and identified by an independent regulator, should be treated as a design problem, and regulation should incentivise platforms to change their systems and processes to reduce it.

Much of the criticism of the inclusion of legal but harmful content within the Bill focuses solely on what harmful speech might be removed by this legislation and ignores the plethora of voices that are already suppressed online due to the often harmful and toxic online environment. If done properly, the inclusion of legal but harmful content within the scope of this legislation could dramatically increase the ability for a wider range of people to exercise their free speech online by increasing the plurality of voices on platforms, especially from minority and persecuted communities. If we are genuinely committed to promoting democratic debate online, preserving the status quo and continuing to exclude these voices is not an option.

## LEGAL BUT HARMFUL IS NOT NEW

Many of those who have opposed the inclusion of legal but harmful content within the Bill have portrayed this as a new and unique ‘threat’ to freedom of speech. In truth it is neither, and the requirement to mitigate the effect of harmful but legal content is already addressed in existing legislation. Broadcast media, for example, has long had legal obligations in this area.

The 2003 Communications Act placed a duty on Ofcom to set standards for the content of programmes, including “that generally accepted standards are applied to the content of television and radio services so as to provide adequate protection for members of the public from the inclusion in such services of offensive and harmful material”. That requirement stemmed from a consensus at the time that broadcasting, by virtue of its universality - a place in virtually every home in the country, and therefore its influence on people’s lives—should abide by certain societal standards. The same could be said now about social media, which is even more ubiquitous and, arguably, more influential, especially for young people.

Similarly, in June 2021, the Carnegie Trust published an extensive response to the Bill, which highlighted how this principle is also already imposed on some user-to-user platforms via the Communications Act:

Note that the Communications Act already imposes on some user-to-user platforms the obligation to “protecting the general

public from videos and audio-visual commercial communications containing relevant harmful material". "Relevant harmful material" includes material containing violence or hatred against a group of persons or a member of a group of persons based on any of the grounds referred to in Article 21 of the Charter of Fundamental Rights of the European Union of 7 December 2000. While some of this material would fall under illegal content, not all the categories are protected by the criminal law. This means that any such types of content would fall to be assessed under 46(3), which might lead to difficulties in the context of lots of low-grade abuse. Arguably, this then constitutes a reduction in the level of protection. The commitments made in the G7 communique about tackling forms of online gendered abuse will be in part delivered by this clause and to set a strong international lead, the clause needs to be made to work.

Clearly, the inclusion of legal but harmful content in the Online Safety Bill is not a fundamentally new principle and as such does not pose a novel threat to freedom of expression as argued by some opponents.

# WHAT IS CLASSED AS “DEMOCRATICALLY IMPORTANT” CONTENT?

One of our major concerns about the draft legislation is that, at present, the vague protections of “democratically important” content could, if not defined thoroughly, actually open up the opportunity for abuse by far-right activists and organisations.

Section 13 of the Draft Online Safety Bill outlines “Duties to protect content of democratic importance”:

(6) For the purposes of this section content is “content of democratic importance”, in relation to a user-to-user service, if— (a) the content is— (i) news publisher content in relation to that service, or (ii) regulated content in relation to that service; and (b) the content is or appears to be specifically intended to contribute to democratic political debate in the United Kingdom or a part or area of the United Kingdom.

The press release that accompanied the publication of the draft Bill clarified that this will include: “content promoting or opposing government policy or a political party ahead of a vote in Parliament, election or referendum, or campaigning on a live political issue.”

While the aim of protecting speech of democratic importance is correct, and one we share at HOPE not hate, it is important that this legislation takes a broad and holistic approach to what is classed as “content of democratic importance”. It is important that any discussion about how this Bill protects democratic speech goes beyond limiting censorship, and includes the promotion of a genuinely pluralistic online space. This demands an analysis of the voices that are so often missing or marginalised online, namely the voices of minority and persecuted communities. We will only create a genuinely democratic online space by broadening out the definition of “democratically important” to include not just content that is often removed, but also content that is missing in the first place. It cannot just protect existing “democratically important” speech, it must also create a safe and pluralistic online space that encourages and empowers diverse and marginalised voices, enabling them to be heard.

However, we also have a serious concern that at present, a lack of clarity around definitions, opens this clause up for abuse by the far right. What happens if content produced by a far-right politician or journalist is also harmful? What is more important in this Bill – reduction of harm caused by hateful online content or protection of “democratically important” speech? What happens when they come into conflict?

At present what will be classified as “democratically important” remains far too vague. The definition at present seems very narrow and imprecise. It seems that speech related to elections will be protected as well as speech related to “live political issues”. Again, how is the latter defined?

The Bill indicates that content will be protected if created by a political party ahead of a vote in Parliament, election or referendum, or campaigning on a live political issue. Will this clause mean that far-right figures who have already been deplatformed for hate speech must be reinstated if they stand in an election? Does this include far-right or even neo-Nazi political parties?

For example, if immigration was deemed a “live political issue” could far-right “migrant hunters” demand that their prejudiced content is protected?

If there were a “grooming gang” case going through the courts, could local far-right activists claim that their anti-Muslim content is protected as it is a “live political issue” in their community?

Under this proposed draft, it could be the case that racist and misogynist content that is legal could be re-uploaded if the content in question was later deemed to be either “democratically important” or a “live political issue”.

# WHAT IS CLASSED AS “JOURNALISTIC CONTENT”?

Much of the research team at HOPE not hate would class themselves as journalists and are members of the National Union of Journalism. As such, we welcome the principle that journalists must be properly protected from undue state interference that could, if done badly, reduce the freedom of the press.

However, we are also very aware that many of the far-right figures we monitor self-define as journalists and will likely seek to exploit these protections to propagate harm online if allowed to do so.

Section 14 of the draft Bill outlines “Duties to protect journalistic content” which includes “a dedicated and expedited complaints procedure available to a person who considers the content to be journalistic content.”

(8) For the purposes of this section content is “journalistic content”, in relation to a user-to-user service, if— (a) the content is— (i) news publisher content in relation to that service, or (ii) regulated content in relation to that service; (b) the content is generated for the purposes of journalism; and (c) the content is UK-linked.

In short, it seems that journalistic content is simply defined as content “generated for the purposes of journalism”.

The press release that accompanied the draft Bill stated that “Articles by recognised news publishers shared on in-scope services will be exempted” and that:

This means they [Category 1 companies] will have to consider the importance of journalism when undertaking content moderation, have a fast-track appeals process for journalists’ removed content, and will be held to account by Ofcom for the arbitrary removal of journalistic content. Citizen journalists’ content will have the same protections as professional journalists’ content.

While it is imperative that this legislation doesn’t unduly curtail the freedom of the press, the vague definition of who is and is not classed as a journalist again opens up the possibility for abuse by the far right.

Some of the most high profile and dangerous far-right figures in the UK, including Stephen Yaxley-Lennon (AKA Tommy Robinson) now class themselves as journalists. There are also far right and conspiracy theory “News Companies” such as Rebel Media and Alex Jones’ InfoWars. These both replicate mainstream news publishers but are used to spread misinformation and discriminatory content. Would this clause mean that their content would receive additional protections?

Similarly to the questions around what is classed as ‘democratically important’ speech, here there seems to be an assumption that journalistic content cannot or does not cause harm. However, what happens when under this proposed draft, it could be the case that racist and misogynist content that is legal could be re-uploaded if the content in question was produced by a journalist? It remains unclear whether it is deemed possible for “journalistic content” to cause harm online? What happens if content produced by a journalist is also harmful? What is more important in this Bill – reduction of harm caused by hateful online content or protection of content by journalists? What happens when they come into conflict?

It also remains unclear whether the bar for removal of harmful content is higher if produced by a journalist and published by a news outlet than if the same opinion/statement is posted by an individual user?

# WILL THIS BILL ADEQUATELY DEAL WITH SMALL BUT EXTREME PLATFORMS?

The Draft Bill states that as soon as reasonably practicable OFCOM must “establish a register of particular categories of regulated services”, splitting them into Category 1, 2A and 2B, each with threshold conditions.

This is extremely important as according to the Government’s response to the White Paper, only Category 1 services “will additionally be required to take action in respect of content or activity on their services which is legal but harmful to adults.” The aim of this is to “mitigate the risk of disproportionate burdens on small businesses.”

However, as we know through our research at HOPE not hate, some of the most dangerous platforms used by the far right to spread hate and organise are smaller platforms they have co-opted, or even small platforms they create themselves (e.g. Bitchute). Will these platforms be classed as Category 1 due to the danger they pose, despite being small and despite this additional regulation being a significant burden on them as a small business?

It is important to understand how the far right use numerous social media platforms

simultaneously. It is incorrect to presume that far-right activists and organisations start on major platforms such as Facebook and Twitter and then move to smaller platforms as they get deplatformed. Often, the far right use platforms of different sizes, with differing amounts of moderation for different purposes. They may use major platforms for their moderate content designed to propagandise and recruit but also use smaller and more laxly moderated platforms to organise and share more extreme content designed to radicalise. They may seek to engage with the general public on major platforms but discuss and plan internally on smaller platforms. Some organisations and people we monitor can be using 5-10 platforms simultaneously, all for slightly different purposes.

The danger is that much of the most dangerous and extreme content is to be found on small platforms. As such, when defining what should be classified as a ‘Category 1 service’ it is necessary to look beyond the size of a platform and also explore the type of harmful content and the user base of that platform.

# IN DEFENCE OF ANONYMITY<sup>1</sup>

In recent months, there have been increasing calls for the legislation to bring an end - or at the very least restrict - anonymity as a means to reducing harmful behaviour online. It is encouraging to see people from across the political spectrum, the government, celebrities, media and the public, demanding change and refusing to accept that abuse is 'just the cost' of being active online.

But it is far from proven whether ending or significantly reducing anonymity online would indeed result in a significant reduction in abuse and harmful behaviour. However, it is likely to significantly reduce the safety of some users online and further marginalise already vulnerable individuals and communities.

HOPE not hate are not completely against some of the tools being discussed to reduce the harm caused by anonymous accounts online, and recognise it can be a problem. However, we strongly believe that any attempt to completely remove anonymity as a right is unacceptable and that we must be careful not to create a two-tier online space between 'verified' and 'anonymous' accounts which could further entrench existing inequalities online.

In addition to whether removing or reducing anonymity would have a significant effect, it would, from HOPE not hate's perspective, likely significantly reduce our, and other investigative journalists', effectiveness when doing research.

People have a right not to be abused online but we also have the right to remain anonymous. We shouldn't give away one right to try and protect the other; we should find solutions to protect both. We should not undermine the rights of citizens to compensate for platforms' bad design decisions that encourage abuse to proliferate.

## WOULD REMOVING ANONYMITY WORK?

Firstly, it's likely that restricting anonymity would not significantly increase our ability to tackle online abuse. While it no doubt emboldens some to behave in a worse fashion, large amounts of the hate and abuse carried out online is done

by *named* people, suggesting this is an issue of ideology and behaviour, not just accountability.

Conversely, requiring identity verification is not needed to ensure accountability. Anonymous accounts can still be blocked, banned or suspended, and many people who post under a pseudonym or have not verified their identity to a platform are still identifiable: law enforcement and NGOs already can and do identify who is behind 'anonymous' accounts which have broken the law to ensure they face ramifications.

## WHY ANONYMITY ONLINE IS IMPORTANT FOR SAFETY

For many people, anonymity online is not antithetical to safety - it *is* safety.

Being able to find information and support, develop opinions and try out different identities without having to declare or prove your identity can be a crucial lifeline. Anonymity allows people who are at risk - which could be related to their sexuality, gender, health, immigration status, or if they are experiencing abuse - to access information, help and support while avoiding questions, persecution or violence.

It is also an essential tool used by civil society - including HOPE not hate when safely researching extremism - plus investigative journalists and whistleblowers to speak truth to power.

## GROUPS FOR WHOM ANONYMITY CAN BE IMPORTANT:

- LGBT+ people
- Undocumented people
- Sex workers
- People seeking health advice and/or end of life care
- Victims of domestic abuse or persecution
- Whistleblowers
- Investigative journalists and researchers
- Political dissidents

<sup>1</sup> Parts of this section were written collaboratively with Ellen Judson who is a Senior Researcher at the Centre for the Analysis of Social Media (CASM) at the cross-party think tank Demos. It was used as a briefing.

## ANONYMITY IS A RIGHT NOT A PRIVILEGE

Anonymity is not just important for people who are particularly vulnerable. It is a right we all have and should protect - particularly as data collected and shared about us online grows exponentially, and people face threats to privacy from governments, corporations, malicious individuals and criminals.

As public discourse plays out online, we should be free to engage without having to hand our identification papers over to a major tech company in Silicon Valley or Beijing.

## WHAT CAN BE DONE TO COMBAT ANONYMOUS ABUSE ONLINE?

The problem with anonymous abuse is not that it's anonymous, but that it's abuse.

We can't look to tech solutions to fix the problem of racism in society, but we should demand tech solutions to problems exacerbated by tech, such as the instantaneous abuse at scale that it facilitates.

The news that the government will designate racist abuse a priority harm under the Online Safety Bill is welcome. But the focus is on post-hoc takedown and user identification, rather than reducing the risk of people experiencing abuse in the first place.

### ■ Give Users More Power Over What They See

One way to maintain people's right to be anonymous online while reducing the harm caused by those who use the cloak of anonymity to cause harm is to give social media users greater control over the content they see online. For example, providing tools that allow users to customise their own networks and limit their interaction with anonymous accounts. This could involve making identity verification an option but not a requirement of engaging on social media.

However, this is only a sticking-plaster solution, and carries risks of its own. It puts the onus on individuals to manage the problem, and doesn't tackle non-anonymous abuse. It does risk, however, that unverified accounts (which are often used by marginalised people who already face barriers to engaging online) are afforded fewer freedoms online, and risks dividing online spaces into separate bubbles of 'anonymous' and 'verified' users. The [government proposals](#) that anonymous accounts have limited functionality, including limited access to end-to-end encryption, would disproportionately affect marginalised and at-risk people.

We propose that any legislation or code of practice that relates to anonymity specifically focus on how platforms are built, designed and run: these could include:

### ■ Stop Algorithms Promoting Harmful and Divisive Content

At present algorithms promote divisive content and 'out-group animosity', rewarding hostility online with virality, while throwaway accounts can be created, deleted and re-created with no curation or oversight. Algorithmic curation systems should be tested and updated to promote socially positive content.

### ■ Encourage Anonymous But Stable Identities Online

Stable identities online can be encouraged by requiring anonymous or verified users to use an account over a period of time, build a network and reputation, engage with others and follow the rules in order to earn more freedoms on the platform.

Greater friction can be introduced when posting, like restricting how brand new accounts can interact with public figures.



HOPE not hate Ltd  
PO Box 61382, London N19 9EQ  
Registered office: Suite 1, 3rd Floor,  
11-12 St. James's Square, London SW1Y 4LB, UK