



**Twitter International
Unlimited Company**

1 Cumberland Place
Fenian Street, Dublin 2,
Ireland
D02 AX07

Registered Number:
503351

Directors:

L. O'Brien,
S. McSweeney
S. Edgett (US)

Mr Andrew Walker
HM Senior Coroner North London
Barnet Coroner's Court
29 Wood Street
Barnet
EN5 4BE

9 February, 2023

**Re: Inquest touching upon the death of Molly Russell: Response to
Regulation 28 Report to Prevent Future Deaths**

Dear Senior Coroner,

1. Thank you for your Regulation 28 report to Prevent Future Deaths (the **Report**) dated 13 October 2022, in which you asked a number of parties, including Twitter International Unlimited Company (formerly Twitter International Company) ('**Twitter**') to respond to concerns arising following the inquest into the death of Molly Russell. We are grateful to you for affording us an extension of time to provide you with our response.
2. We would like to begin by extending our deepest sympathies to Molly's family and friends for the loss they have suffered.
3. The purpose of this letter is to set out the steps taken, or intended to be taken, by Twitter in respect of the six matters of concern detailed in your Report. As you will be aware, Twitter was not given Interested Person status at the inquest and so in preparing this response we have not been able to consider the evidence made available to the inquest. Notwithstanding, we have carefully considered the recommendations set out in your Report in line with our ongoing commitment to ensuring our platform is a safe space for all users. We have also noted that many of your concerns are currently subject to Parliamentary debate in relation to the draft

Online Safety Bill. Twitter welcomes the enactment of the Bill and is hopeful that it will create an appropriate framework for balancing the complex challenge of content regulation with the benefits of social media, balancing respective freedoms and rights of users fairly.

4. **Concern 1: separate platforms for adults and children; and Concern 3: controls to ensure content is age specific**

4.1. In accordance with regulatory requirements in the US, UK and Europe, Twitter requires its users to be aged 13 or over. It does not currently have a separate platform for users aged between 13 and 16, or those under 18. Instead, the platform is designed to be a different experience for younger users, while all users are provided with tools to tailor the types of content they are presented with to suit their circumstances. It is worth noting that the average age of a Twitter user is older than other social media platforms. Research carried out by Comscore reported that as of December 2022, 98% of Twitter users are over the age of 18. Notwithstanding, Twitter is designed to be age appropriate for teenagers from age 13 years and up.

4.2. There are a number of challenges to any social media platform in creating a separate platform for teenage users, while the benefits of segregated platforms are not clear. A proportion of teenagers will always discuss their emotions and mental health challenges on social media. Sharing a platform with adults provides an opportunity for supervision and support to be provided to teenagers, in circumstances where teenagers segregated on a platform may not be as well equipped to respond appropriately to such content.

4.3. Rather than segregating platforms, Twitter has designed its platform to provide a different experience for younger users as well as deploying a number of safety features in order to keep all users safe. By way of example:

4.3.1. Age restricted content – Twitter automatically restricts users who are under 18, or who do not include a birth date on their profile, from viewing sensitive media content (as set out in our sensitive media policy)¹. In addition, a different approach to advertising is taken for users who are either under 18 or who do not include a birth date on their profile. Twitter prohibits marketing or advertising of a number of products and services to minors, including alcohol, weapons, weight loss products, health supplements, gambling products, sexual products and services, permanent cosmetics and other forms of body branding². These age restrictions are in addition to complete bans on advertising certain products on Twitter,

¹ <https://help.twitter.com/en/rules-and-policies/notices-on-twitter>

² <https://business.twitter.com/en/help/ads-policies/ads-content-policies/prohibited-content-for-minors.html>

including any advertising of controlled substances, tobacco and projectiles.

- 4.3.2. Safe Search – users of the Twitter platform have control over what they can see in search results through selecting the Safe Search mode. Safe Search is automatically enabled for anyone with a birth date under 18 years of age. Once enabled, these filters are designed to exclude from search results any potentially sensitive content (such as content which is excessively gory, violent, or of a graphic sexual nature)³ along with accounts a user has muted or blocked (for whatever reason).
- 4.3.3. Sensitive Tweet Warnings – Twitter’s sensitive media policy prohibits users from including graphic content or adult nudity and sexual behaviour within areas that are highly visible on Twitter, including in live video, profile, header, List banner images, or Community cover photos. If a user shares this content on Twitter, the policy requires the user to mark their entire account as sensitive or to add sensitive content warnings to individual photos or videos. Doing so places an interstitial warning message on images or videos they post which contain sensitive media. Twitter may also place an interstitial warning message on some forms of sensitive media. An interstitial warning alerts a user that a Tweet contains sensitive content such as nudity, violence or sexual content and means other users can only see the media if they actively click to "show" the Tweet; it cannot be viewed by accident.
- 4.3.4. Controlling replies – users can choose who will be able to reply to their Tweets when posted. The default position is that everyone can reply but options are available to turn off all replies or only allow the accounts mentioned in the Tweet to reply. A user can also change who can reply to their Tweets, or turn off replies, after the Tweet has been posted.
- 4.3.5. Protected accounts – when an adult user signs up for Twitter, they can choose to keep their Tweets public or to protect them so that only approved followers can see and interact with them⁴. By contrast, when a user signs up for Twitter with a date of birth under 18 years of age, the account is automatically defaulted to protected mode.
- 4.3.6. Account filters – users can filter the types of accounts they see in their notifications timeline. This feature allows users to mute notifications from certain categories of users, such as those with accounts who have not confirmed their phone number or email address, new accounts, accounts

³ <https://help.twitter.com/en/rules-and-policies/media-policy>

⁴ <https://help.twitter.com/en/safety-and-security/public-and-protected-tweets>

who have a default profile photo, accounts that the user does not follow or accounts that do not follow the user⁵.

4.3.7. Block and mute – users can block accounts instantly if they do not want that account to see their Tweets and/or the user does not want to see the account's Tweets. Users can also mute an account if they don't want to see their Tweets, but don't want to unfollow the account. Particular words, conversations, phrases, emojis and hashtags can also be muted to ensure those words or phrases do not appear on the user's timeline.

5. **Concern 2: age verification when signing up to the platform**

5.1. Twitter is committed to protecting child safety online and has launched a range of age assurance measures to seek to ensure that only users aged 13 and over are permitted to access the Twitter platform.

5.2. As previously noted, Twitter requires its users to be at least 13 years old in order to create an account. Twitter approaches the challenge of age assurance by combining self-declaration (i.e. users providing their date of birth) with additional technical measures (as described in the ICO's Age-Appropriate Design Code⁶) which together aim to ensure that the account holder's self-declared age is genuine and that appropriate controls are in place to protect teenagers.

5.3. Twitter first collects the user's age through the neutral presentation of a date of birth prompt. Once a date of birth is entered, Twitter then determines the user's age. At this stage, new users are informed that Twitter uses their age to customise their experience, including advertising, and provides options as to the visibility of the user's date of birth to others.

5.4. Users who enter a date of birth that indicates they are under the age of 13 are not permitted to go any further in the account opening process. There is an account restoration appeals process for those who erroneously enter the wrong date of birth and are not permitted to proceed with account opening or who have their account off-boarded as a result of an indication of being under 13. As part of the account restoration appeals process, the user is required to provide ID documentation proving that they are over the age of 13. These appeals are subject to human review. If Twitter cannot verify the user is over the age of 13, the account is not restored. (This appeals process is often used by business accounts who enter the date of incorporation, rather than children seeking to attempt to gain access to Twitter. Where the account is registered to a legal

⁵ https://twitter.com/settings/notifications/advanced_filters

⁶ See in Chapter 3 of the AADC 'How can we establish age with an appropriate level of certainty' available at <https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-a-code-of-practice-for-online-services/>

person (i.e. a company), evidence would need to be provided to show that the account is being used for business purposes.)

5.5. Users who enter a date of birth that indicates they are over 13 but under 18 are prevented from seeing sensitive content, such as adult content on any surfaces (e.g. their timeline or search results) in line with Twitter's sensitive media policy and the automatic application of 'Safe Search' for such users. Any sensitive content contained in the account holder's page will be obscured by a sensitivity screen, in line with the policies identified at paragraph 4.3.3 above.

5.6. Users are also able to report accounts which they believe are operated by someone who is underage and Twitter will take action if appropriate.

5.7. In respect of advertising, users who have not registered a date of birth on their profile (for example, because they opened their account before providing a date of birth was required), will be asked to enter their date of birth in order to follow the accounts of certain brands. Twitter prohibits marketing or advertising of a number of products and services to minors, such as alcohol. If the user is a minor, these types of ads will not be served to them, as explained in further detail in paragraph 4.3.1 and 6.2.

5.8. In addition to the measures above, Twitter has been working with experts to research further age assurance measures that incorporate 'privacy by design' principles (required by the GDPR) and work in a global context. These measures also need to account for the importance of online anonymity for minorities and disadvantaged communities around the world and the use of Twitter as a platform for whistle-blowers and human rights advocates.

5.9. There are currently a range of projects which are being actively examined by Twitter with these considerations in mind, focused on the best interests of children.

6. **Concern 4: algorithms used to provide content together with adverts**

6.1. As you may be aware, the majority of online services use algorithms in some form to suggest relevant content to users, which helps improve the usability and accessibility of online services.

6.2. Twitter uses algorithms to help provide content to users. The main feed on Twitter is sub-divided between a 'Following' tab (which only shows Tweets posted or Re-Tweeted by accounts a user is following) and a 'For You' tab (which suggests more Tweets from accounts and topics a user follows as well as recommended Tweets). Users may also see content such as

Promoted Tweets or Re-Tweets in their timeline⁷. Neither tab permits sensitive content or inappropriate advertising to be surfaced for users under the age of 18. Twitter's policies and enforcement measures seek to reduce the risk that illegal or potentially harmful content could be shown to users.

- 6.3. Twitter's Suicide and Self Harm policy prohibits users from promoting or encouraging suicide or self-harm⁸. If this policy is violated (e.g. the user shares content which intentionally encourages others to harm themselves, asks others to encourage the user to harm themselves or shares detailed information or instructions relating to self-harm or suicide), Twitter actions the content so it is no longer visible publicly and requires the user to remove the content. The user will be unable to Tweet again or interact in any way on the platform until they do so. If a user continues to violate Twitter's Suicide and Self Harm policy, or if an account appears dedicated to promoting or encouraging self-harm or suicide, the account will be permanently suspended. In addition to content removal, Twitter also marks hyperlinks as unsafe; for example, where a link may be seeking to spread instructional material⁹.
- 6.4. Twitter's Suicide and Self Harm policy was developed after consulting extensively with experts. The policy does not prevent people who have engaged in self-harm or experienced suicidal thoughts from sharing their personal experiences and using the platform for seeking support. Experts believe that removing posts of this nature risks not only stigmatising mental health challenges but also removes an opportunity for intervention by the friends and family of a user.
- 6.5. Twitter has also launched a new product called '#ThereIsHelp' in the UK¹⁰. This means a prompt with a link to the Samaritans charity will appear when a user searches for words related to suicide or self-harm. On the mobile app, the mode in which the majority of users access Twitter, the prompt takes up almost half the screen.
- 6.6. During the last reporting period, there was a substantial increase in the volume of accounts suspended (18% increase), and content removed (23% increase) under Twitter's 'Promoting suicide or self-harm' policy. 408,143 accounts were actioned in total. We attribute this increase to our continued investment in identifying violative content at scale. As a business we are determined to continue improving in this area. To improve transparency, we regularly publish data around Twitter's enforcement of its policies¹¹.

⁷ <https://help.twitter.com/en/using-twitter/twitter-timeline>

⁸ <https://help.twitter.com/en/rules-and-policies/glorifying-self-harm>

⁹ <https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links>

¹⁰ https://blog.twitter.com/en_us/topics/company/2018/wspd2018

¹¹ <https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jul-dec>

7. Concern 5: Parental access and control over the material being viewed and Concern 6: linking to parental accounts for monitoring

7.1. As previously stated, users under 18 make up a very small minority of all Twitter users in the UK. Notwithstanding this, our Trust and Safety Team is dedicated to advocating for the safety of its users and protecting their rights, and therefore engages with experts to ensure Twitter offers the most appropriate solutions to parents with children using Twitter. In collaboration with Internet Matters (an organisation launched with the specific intention of supporting parents and carers to navigate the digital landscape), Twitter has developed a parental controls guide, which provides step-by-step instructions for parents to manage their child's account¹².

7.2. These instructions allow parents to protect their child's Tweets (as described at paragraph 4.3.5 above) and prevent children from receiving abusive or inappropriate content. It also gives the parent control over who can contact their child and what personal data is shared. The controls also allow parents to limit who can see their child's Tweets, who can contact them and who can tag them.

7.3. As explained above, users can curate the types of content they see to match their interests and hide Tweets that contain sensitive content. In addition, Twitter introduced 'Safety Mode' in September 2021¹³, which allows users to temporarily block accounts for using potentially harmful language or sending repetitive and uninvited replies or mentions.

8. Concluding remarks

8.1. We hope that this response provides you with a helpful explanation of the steps Twitter has already taken in relation to your concerns. Twitter does not underestimate the challenge in this area. We are committed as an organisation to working with experts, regulators, government and others in the sector to ensure that online services are as safe as they can be for its users and in particular those under the age of 18.

Yours Sincerely,

Twitter International Unlimited Company

¹² <https://www.internetmatters.org/parental-controls/social-media/twitter/>

¹³ https://blog.twitter.com/en_us/topics/product/2021/introducing-safety-mode