

Know Your Digital Rights

How you can hold social media platforms accountable for online antisemitism

Protected Characteristics

All mainstream social media platforms have community standards or hate speech policies which ban discrimination and hate speech targeting people on the basis of what are defined as “**protected characteristics**”. **The broad spectrum of protected characteristics across platforms include: race, ethnicity, nationality, national origin, citizenship, statelessness, religious affiliation, sex, sexual orientation, gender, gender identity, age, disability, disease, and caste.** As a Jewish user, this means that social media companies have promised you protection from online hate hosted on these platforms – whether it is targeted harassment against you specifically, or generally spreading fear and harmful misinformation about the Jewish people as a group.

Reporting Online Hate

Reporting offensive content on each platform is different and can be frustrating. Based on our experience, typically the first time a piece of content is reported by a user to a social media platform, their artificial intelligence (AI) reviews the content to see if any of the images or words ‘trigger’ automated programming to remove the content. For example, an image of a Swastika has a high likelihood of being removed across most platforms most times that it is reported. However, the current AI models often miss certain classic images of antisemitism, nuanced text alluding to antisemitic conspiracy theories, and modern forms of antisemitism, like virulent anti-Israel sentiment.

If you submit a report and the platform says that it does not violate their standards, don’t be frustrated. **Typically, you can report the content again.** A **second report** by the same user has a higher likelihood of being **reviewed by a human instead of AI.** Unfortunately, platforms often out-source human-review to content review teams that either do not have the capacity or the necessary education to address the volume or nuance of online antisemitic content. [Below](#), we have featured a how to guide, with some basic visual guidance on how to report Jew-hatred on your preferred platform.

The Loophole

Despite the digital policies and standards that platforms created to ensure a harassment-free zone for protected groups, we know that antisemitic content remains online despite being reported at overwhelming rates. How is this possible? How can social media platforms continue to host hate, conspiracies, along with other harmful messaging, some even with direct connections to offline violence?

[Section 230 of the 1996 Communications Decency Act](#) protect social media platforms from legal liability relating to harmful content posted on their platforms by third parties (including, hate group and regular users). While the same law also empowers the social media companies to monitor and remove any harmful content at their discretion, they are not legally required to do so. Since the spread of online hate and other unforeseen social harms propelled by widespread social media use, especially by children and teens, have come to light both Republican and Democrats have increasingly criticized the Section 230 loophole and have proposed bi-partisan changes to law—[over 20 times](#) in the current 117th Congress alone!¹

Meanwhile, the European Union's [Digital Services Act](#) (DSA), passed in July 2022, will start to enter into force this November 16, 2022 and be fully rolled out across Europe in February 2024. The DSA includes [core requirements](#) of large social media platforms & search engines, including increased transparency & reporting, compliance by design to easily remove illegal content, including hate speech & systematic user complaints management. The DSA empowers the EU to fine social media platforms up to 6% of their annual revenue if they fail to comply!

The Accountability Crisis

The sad truth is that today social media platforms still rely on their users to report violating content to them before responding. They are not actively seeking out and removing harmful content themselves without concerted public pressure campaigns, research from civil society or regulation. That's like a police department saying that they are only going to enforce laws when they receive reports from the public. We know that technology is available that enables platforms to identify and remove content, and that they do so regularly where there are **legal** consequences for leaving that content online. For example, the best technology in content moderation is applied today to things such as copyright violation and child pornography. Content that violates either of these policies is swiftly removed with severe ramifications for the original content creator. Imagine if the same stringency was applied to flagging hate content.

¹ Chart tracking the proposed changes to Section 230 and relevant sponsoring Congressmen. Compile by Alan Kyle, Mast

While the legal landscape for social media platforms may slowly change, CyberWell is collecting, vetting and reporting data with a unique lens of **digital policy compliance** and sharing that data with the public so that, together, we can change our digital future for the better.

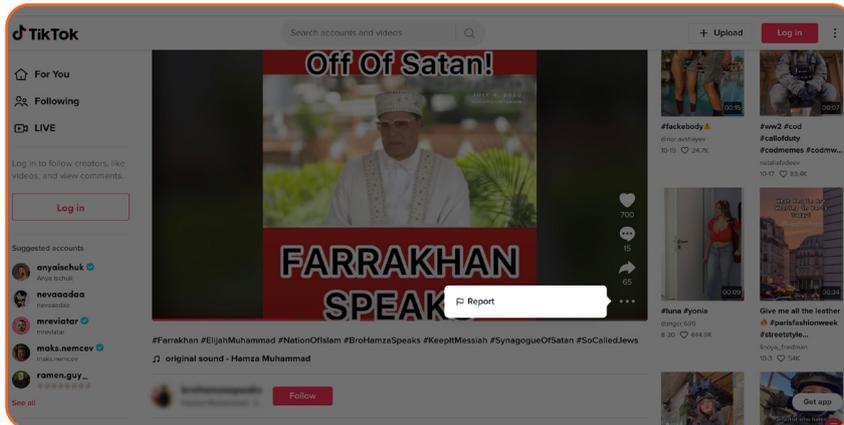
What You Can Do

1. **Call out antisemitism online where you see it – remember, most users are passive, meaning they only read the post and the comments section, but do not engage.** If antisemitic tropes do not get labeled as purveying hate, offense and fear, the passive reader might not register that the content they are viewing is antisemitic.
2. **Support users who openly express their Jewish identity** – It is not easy online for a Jew, especially those who are proudly producing visibly Jewish content. Show your support in the comments section, messages and likes when you can!
3. **Take a proactive stance and report online Jew-hatred via CyberWell's [interactive data platform](#).**
4. **Crowdsource report the online antisemitism you experience to CyberWell** – we will do the follow up work to ensure that the platforms take down this content!

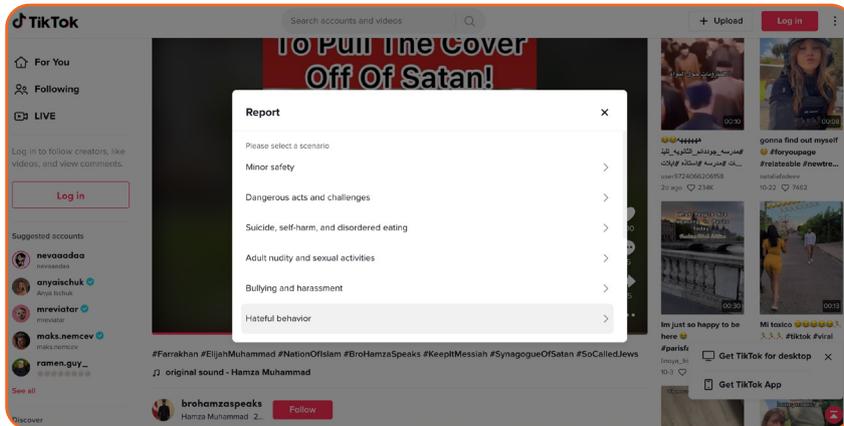
Don't forget – **this is a numbers game**. The more reports a post receives, the greater the likelihood that the platform will remove that content. Do not get discouraged by the reporting processes or responses that a clearly offensive post does not violate a platform's policy. The more we apply consistent, methodical pressure on the social media platforms, the greater our voices become amplified. **We know that social media platforms can do better and, together, we can ensure that antisemitism is unacceptable online as it is offline.**

Reporting Online Antisemitism

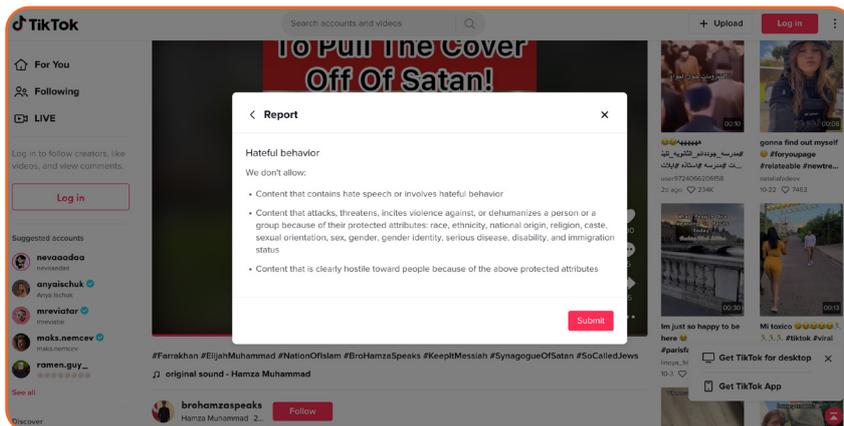
Step-By-Step Guide



Press on the ●●● and Report



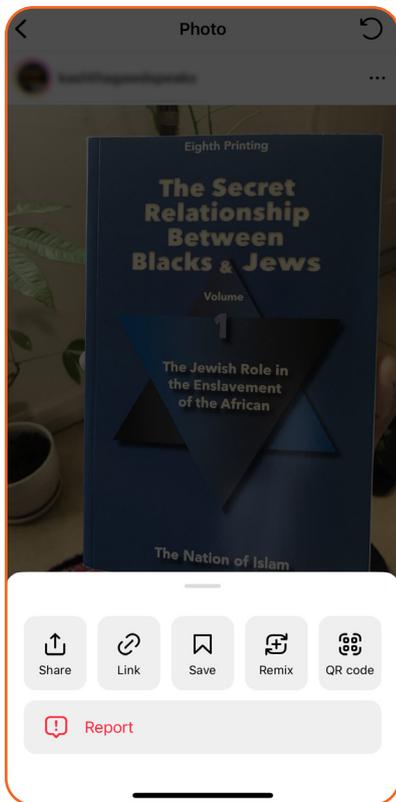
Select Violation



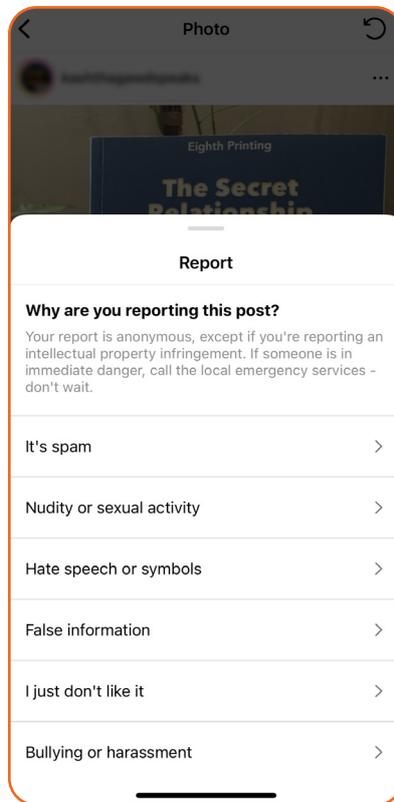
Submit!

Reporting Online Antisemitism

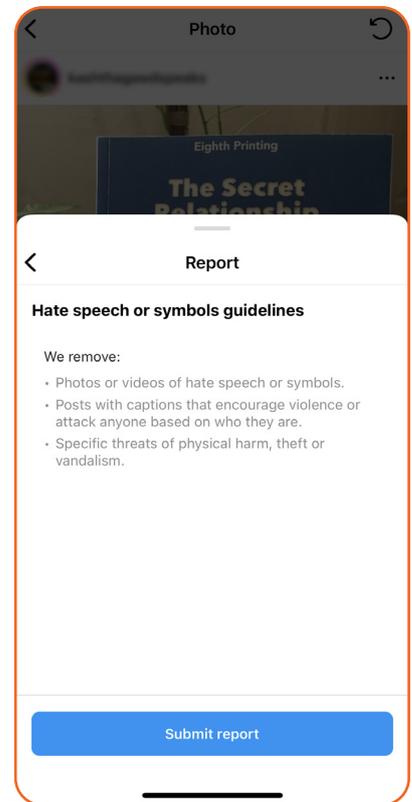
Step-By-Step Guide



Press on the ●●● and Report



Select Violation

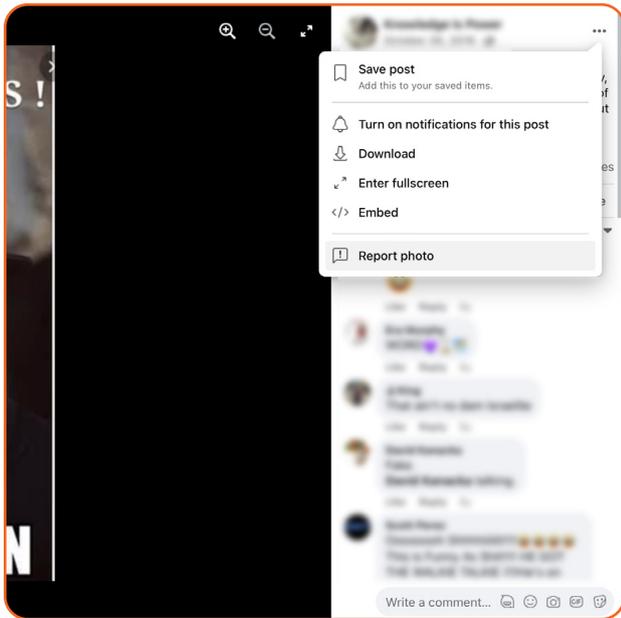


Submit!

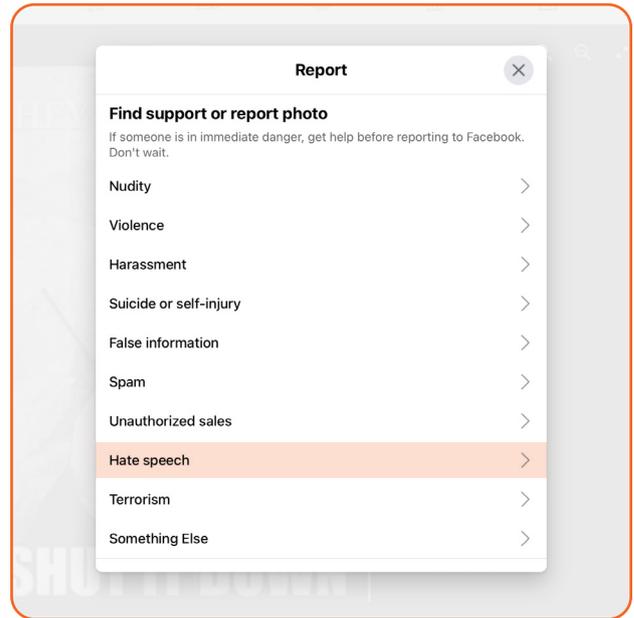
Reporting Online Antisemitism

Step-By-Step Guide

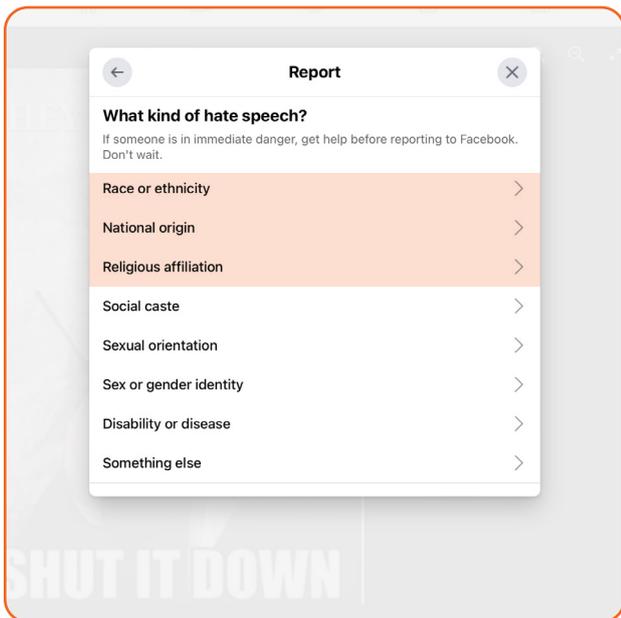
facebook



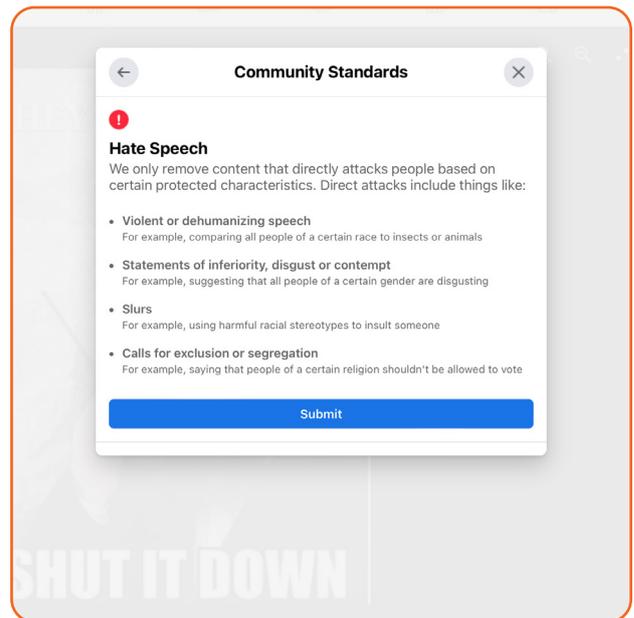
Press on the ●●● and Report



Select Violation



What Kind of Hate Speech? Depending on the context, Religious Affiliation, National Origin, Race or Ethnicity



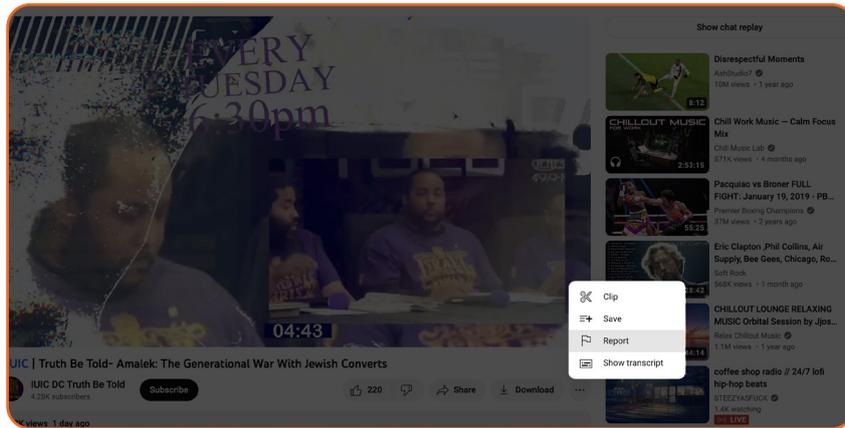
Submit!

Reporting Online Antisemitism

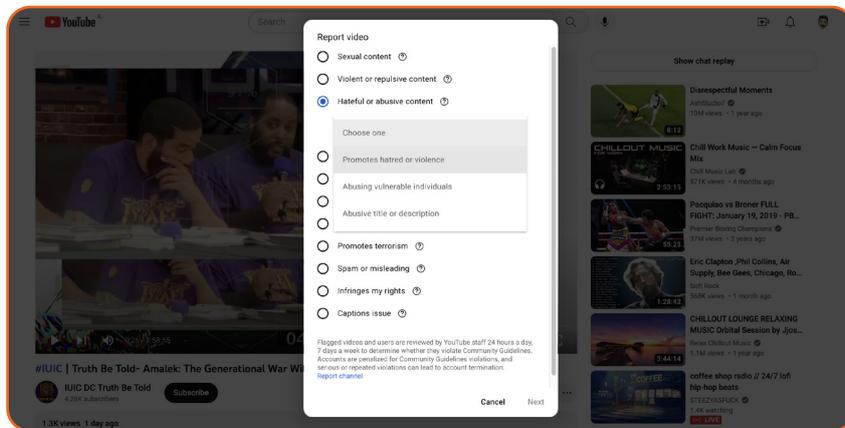
Step-By-Step Guide



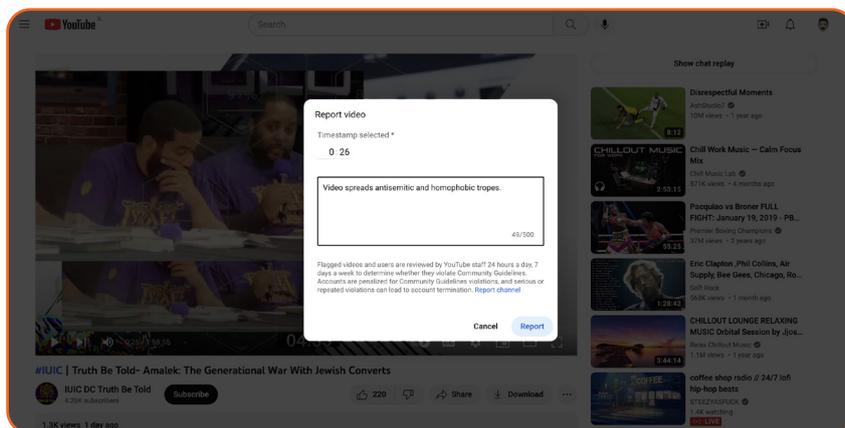
NOTE: Best to pause the video where its antisemitic before making your report



Press on the ●●● and Report



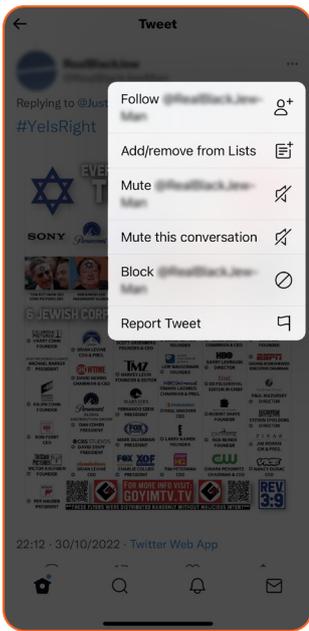
Select Violation



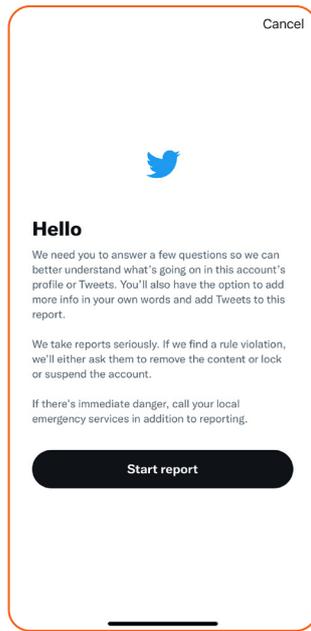
Add more detail (optional) & Submit!

Reporting Online Antisemitism

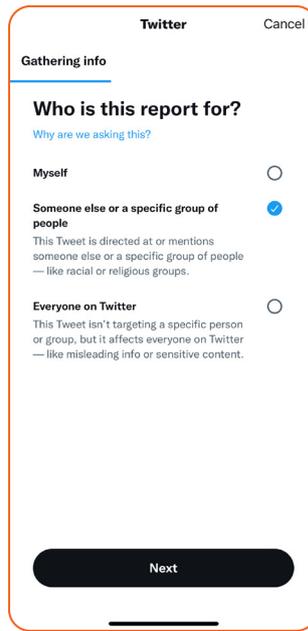
Step-By-Step Guide



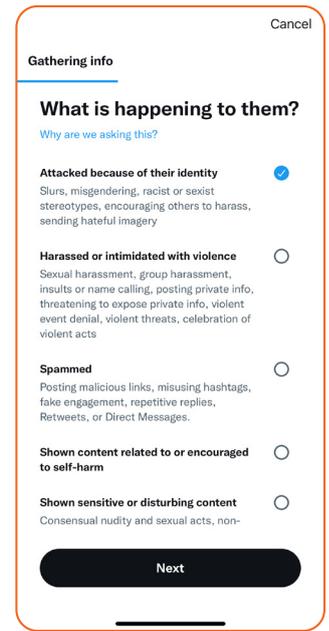
press on the ●●● and Report tweet



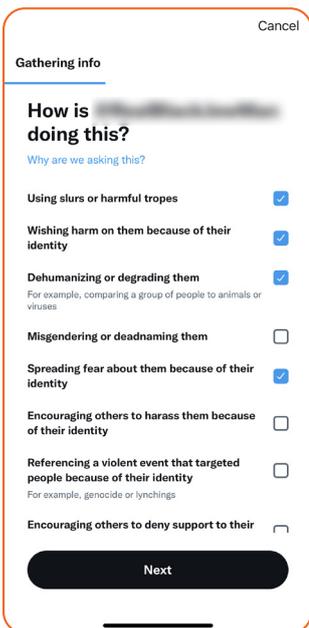
Start the report



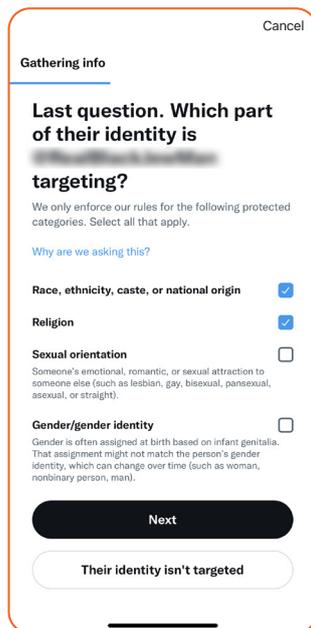
Who is this report for? Depending on the context, Myself or a specific group



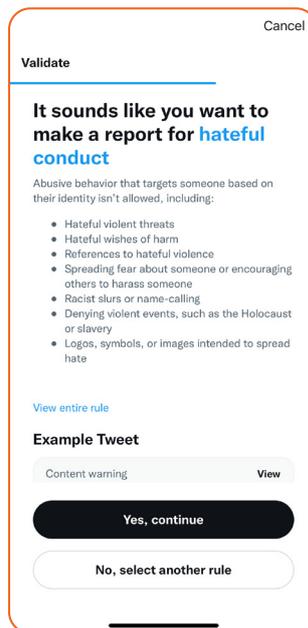
What is happening to them? Typically, Attacked because of their identity



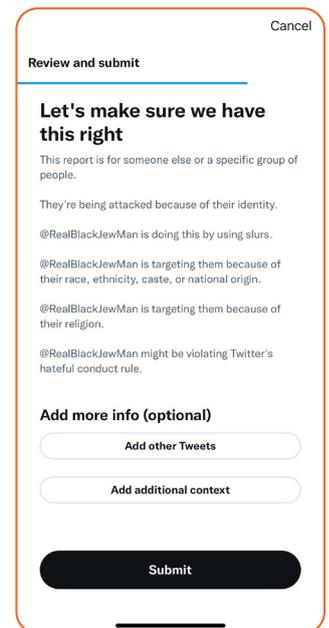
Select how the account is spreading antisemitism



Select which part is being targeted



Validate the selected policy that is being violated



Review & Submit Feel free to add more Tweets that are also antisemitic!