

Misinformation

Policy Rationale

Misinformation is different from other types of speech addressed in our Community Standards because there is no way to articulate a comprehensive list of what is prohibited. With graphic violence or hateful conduct, for instance, our policies specify the content we prohibit, and even persons who disagree with those policies can follow them. With misinformation, however, we cannot provide such a line. The world is changing constantly, and what is true one minute may not be true the next minute. People also have different levels of information about the world around them, and may believe something is true when it is not. A policy that simply prohibits “misinformation” would not provide useful notice to the people who use our services and would be unenforceable, as we don’t have perfect access to information.

Instead, our policies articulate different categories of misinformation and try to provide clear guidance about how we treat that speech when we see it. For each category, our approach reflects our attempt to balance our values of expression, safety, dignity, authenticity, and privacy.

We remove misinformation where it is likely to directly contribute to the risk of imminent physical harm. We also remove content that is likely to directly contribute to interference with the functioning of political processes. In determining what constitutes misinformation in these categories, we partner with independent experts who possess knowledge and expertise to assess the truth of the content and whether it is likely to directly contribute to the risk of imminent harm. This includes, for instance, partnering with human rights organizations with a presence on the ground in a country to determine the truth of a rumor about civil conflict.

For all other misinformation, we focus on reducing its prevalence or creating an environment that fosters a productive dialogue. We know that people often use misinformation in harmless ways, such as to exaggerate a point (“This team has the worst record in the history of the sport!”) or in humor or satire (“My husband just won Husband of the Year.”) They also may share their experience through stories that contain inaccuracies. In some cases, people share deeply-held personal opinions that others consider false or share information that they believe to be true but others consider incomplete or misleading.

Recognizing how common such speech is, we focus on providing users with helpful information when there is potentially misleading or confusing content. As part of that effort, outside of the United States we partner with third-party fact checking organizations in many parts of the world to review and rate the accuracy of the most viral content on our platforms (see [here](#) and [here](#) for our approach in the United States). Beginning in the United States, the Community Notes program lets people add more context to Facebook, Instagram and Threads posts that are potentially misleading or confusing (see [here](#)). We also provide resources to increase media and digital

literacy so people can decide what to read, trust, and share themselves. We require people to disclose, using our AI-disclosure tool, whenever they post organic content with photorealistic video or realistic-sounding audio that was digitally created or altered, and we may apply penalties if they fail to do so. We may also add a label to certain digitally created or altered content that creates a particularly high risk of misleading people on a matter of public importance.

Finally, we prohibit content and behavior in other areas that often overlap with the spread of misinformation. For example, our Community Standards prohibit [fake accounts](#), [fraud](#), and [coordinated inauthentic behavior](#).

As online and offline environments change and evolve, we will continue to evolve these policies. Accounts that repeatedly share the misinformation listed below may, in addition to having their content enforced on in accordance with this policy, receive decreased distribution, limitations on their ability to advertise, or be removed from our platforms. Additional information on what happens when Meta removes content can be found [here](#).

Guidelines

We remove the following types of misinformation:

I. Physical Harm or Violence

We remove misinformation or unverifiable rumors that expert partners have determined are likely to directly contribute to a risk of imminent violence or physical harm to people. We define misinformation as content with a claim that is determined to be false by an authoritative third party. We define an unverifiable rumor as a claim whose source expert partners confirm is extremely hard or impossible to trace, for which authoritative sources are absent, where there is not enough specificity for the claim to be debunked, or where the claim is too incredulous or too irrational to be believed.

We know that sometimes misinformation that might appear benign could, in a specific context, contribute to a risk of offline harm, including threats of violence that could contribute to a heightened risk of death, serious injury, or other physical harm. We work with a global network of non-governmental organizations (NGOs), not-for-profit organizations, humanitarian organizations, and international organizations that have expertise in these local dynamics.

In countries experiencing a heightened risk of societal violence, we work proactively with local partners to understand which false claims may directly contribute to a risk of imminent physical harm. We then work to identify and remove content making those claims on our platform. For example, in consultation with local experts, we may remove out-of-context media falsely claiming to depict acts of violence, victims or perpetrators of violence, weapons, or military hardware.

II. Harmful Health Misinformation

We consult with leading health organizations to identify health misinformation likely to directly contribute to imminent harm to public health and safety. The harmful health misinformation that we remove includes the following:

Misinformation about vaccines. We remove misinformation primarily about vaccines when public health authorities conclude that the information is false and likely to directly contribute to imminent vaccine refusals. They include:

- Vaccines cause autism (Ex: “Increased vaccinations are why so many kids have autism these days.”)
- Vaccines cause Sudden Infant Death Syndrome (Ex: “Don’t you know vaccines cause SIDS?”)
- Vaccines cause the disease against which they are meant to protect, or cause the person receiving the vaccine to be more likely to get the disease (Ex: “Taking a vaccine actually makes you more likely to get the disease since there’s a strain of the disease inside. Beware!”)
- Vaccines or their ingredients are deadly, toxic, poisonous, harmful, or dangerous (Ex: “Sure, you can take vaccines, if you don’t mind putting poison in your body.”)
- Natural immunity is safer than vaccine-acquired immunity (Ex: “It’s safest to just get the disease rather than the vaccine.”)
- It is dangerous to get several vaccines in a short period of time, even if that timing is medically recommended (Ex: “Never take more than one vaccine at the same time, that is dangerous - I don’t care what your doctor tells you!”)
- Vaccines are not effective at preventing the disease against which they purport to protect. However, for the COVID-19, flu, and malaria vaccines, we do not remove claims that those vaccines are not effective in preventing someone from contracting those viruses. (Ex’s: Remove – “The polio vaccine doesn’t do anything to stop you from getting the disease”; Remove – “Vaccines actually don’t do anything to stop you from getting diseases”; Allow – “The vaccine doesn’t stop you from getting COVID-19, that’s why you still need to socially distance and wear a mask when you’re around others.”)
- Acquiring measles cannot cause death (requires additional information and/or context) (Ex: “Don’t worry about whether you get measles, it can’t be fatal.”)
- Vitamin C is as effective as vaccines in preventing diseases for which vaccines exist.

Misinformation about health during public health emergencies. We remove misinformation during public health emergencies when public health authorities conclude that the information is false and likely to directly contribute to the risk of imminent physical harm, including by contributing to the risk of individuals getting or spreading a harmful disease or refusing an associated vaccine. We identify public health emergencies in partnership with global and local health authorities.

Promoting or advocating for harmful miracle cures for health issues. These include treatments where the recommended application, in a health context, is likely to directly contribute to the risk of serious injury or death, and the treatment has no legitimate health use (ex: bleach, disinfectant, black salve, caustic soda).

III. Voter or Census Interference

In an effort to promote election and census integrity, we remove misinformation that is likely to directly contribute to a risk of interference with people's ability to participate in those processes. This includes the following:

- Misinformation about the dates, locations, times, and methods for voting, voter registration, or census participation.
- Misinformation about who can vote, qualifications for voting, whether a vote will be counted, and what information or materials must be provided in order to vote.
- Misinformation about whether a candidate is running or not.
- Misinformation about who can participate in the census and what information or materials must be provided in order to participate.
- Misinformation about government involvement in the census, including, where applicable, that an individual's census information will be shared with another (non-census) government agency.
- False or unverified claims that the U.S. Immigration and Customs Enforcement (ICE) is at a voting location.
- Explicit false claims that people will be infected by COVID-19 (or another communicable disease) if they participate in the voting process.
- False claims about current conditions at a U.S. voting location that would make it impossible to vote, as verified by an election authority.

We have additional policies intended to cover calls for violence, the promotion of illegal participation, and calls for coordinated interference in elections, which are represented in other [sections](#) of our Community Standards.

For the following content, we include an informative label:

Manipulated Media

Media can be edited in a variety of ways. In many cases, these changes are benign, such as content being cropped or shortened for artistic reasons or music being added. In other cases, the manipulation is not apparent and could mislead.

Content Digitally Created or Altered that May Mislead. For content that does not otherwise violate the Community Standards, we may place an informative label on the face of content – or reject content submitted as an advertisement – when the content is a photorealistic image or video, or realistic sounding audio, that was digitally created or altered and creates a particularly high risk of materially deceiving the public on a matter of public importance.