

Managing content responsibly

At YouTube, it is our responsibility to manage what's on the platform so that people can access authoritative information while being protected against problematic content. We are here to explain the three guiding principles – Remove, Reduce, and Raise – that form the basis of our approach to making YouTube safer.



1. Remove content that violates our policies

Developing policies for a global platform

Our policies are designed to make YouTube safer and outline what content is allowed and what is not acceptable to post. Developed in partnership with YouTube creators and external policy experts, they're effective, well-informed, and apply globally.

A few examples of the areas our policies cover are:

- Violent Extremism
- Hate and Harassment
- Child Safety
- Spam and Deceptive Practices

Removing content before it's widely viewed

We use a combination of humans and machines to detect, review, and remove content that violates our policies.

Detecting potentially harmful content

Machines: Flag content at scale

Users: Use a flagging feature

YouTube Trusted Flagger program:

External experts who flag content with high precision.



Reviewing flagged content

Flagged content is reviewed by trained human reviewers to assess whether the content does indeed violate our policies, and to protect content that has an educational, documentary, scientific, or artistic purpose.



Taking action on harmful content

After careful evaluation, human reviewers make decisions to either remove content that violates our policies or age-restrict content that may not be appropriate for all audiences.

Sharing our progress on removing harmful content

Our quarterly **Community Guidelines Enforcement Report** gives insight into the volume of flags and actions taken on content that violates our policies. Removals in Q1 of 2021 included:

9.6M

Videos removed

2.2M

Channels removed

1.0B

Comments removed



2. Reduce the spread of harmful misinformation and borderline content

Borderline content represents a fraction of 1% of what's watched on YouTube. That said, even a fraction of a percent is too much. So we have reduced recommendations of borderline content or videos that could misinform people in harmful ways.

Harmful misinformation

Content that promotes unsubstantiated conspiracy theories, or inaccurate information - whether deliberate (disinformation) or not (misinformation). External human evaluators use public guidelines to determine harmful misinformation.

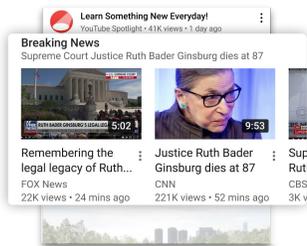
Borderline content

Content that comes close to — but doesn't quite cross the line of — violating our Community Guidelines.

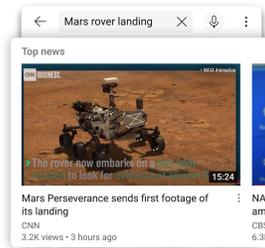
3. Raise up authoritative information

Making authoritative sources readily available

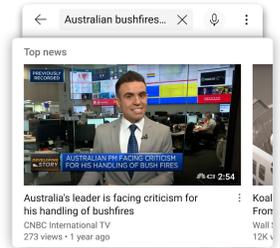
In areas where veracity and credibility are key, we work hard to ensure that our systems prioritize surfacing authoritative content from trusted sources.



Breaking news on the homepage appears during breaking news moments



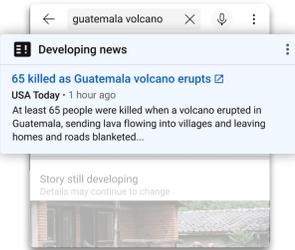
Top news in search results appears based on viewer search query



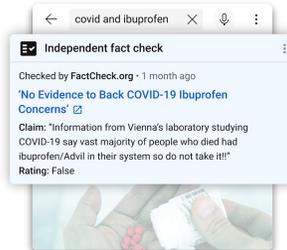
Authoritative sources in search results and "up next" panels get prioritized for newsworthy events

Providing context to help people make their own decision

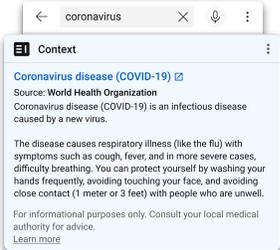
Information panels provide additional context alongside certain search results and videos to help people make more informed decisions.



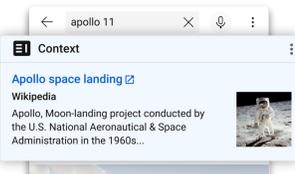
Developing news in search results link to text based news articles from authoritative sources



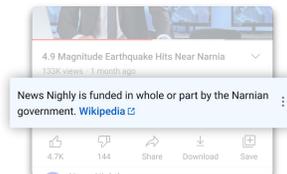
Fact check in search results appear alongside recently fact checked claims by third-parties



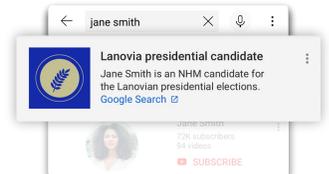
Health panel provides detailed information, including symptoms and prevention



Topics prone to misinformation surface content from third-party sources to provide additional context



Publisher funding is displayed for government or publicly funded publishers



Election candidate search results display text-based information during election periods