## Starter: Check out this thread from @TwitterSafety that walks through Twitter's principles & recent developments

### COVID-19

As the global community faces the COVID-19 pandemic together, Twitter is helping people find reliable information, connect with others, and follow what's happening in real time. You can find all of our updates here. As regards misleading information specifically, these blogs are instructive:

- Updating our approach to misleading information
- Broadening our guidance on unverified claims
- Broadening our definition of "harm"

### Twitter Transparency Report

For the past eight years, our biannual Twitter Transparency Report has highlighted trends in requests made to Twitter from around the globe. Over time, we have significantly expanded the information we disclose by adding metrics on platform manipulation, Twitter Rules enforcement, and our proactive efforts to eradicate terrorist content. Here you can read the blog accompanying the most recent edition of the TTR.

### Transparency – Information Operations

In October 2018, we published the first comprehensive archive of Tweets and media associated with known state-backed information operations on Twitter. Since its launch, thousands of researchers from across the globe have downloaded datasets, using our archive to conduct their own investigations and to share their insights and independent analysis. Researchers can apply for access here. You can read this blog which accompanied the most recent release earlier this month.

### Platform Manipulation – Bot or not?

Going back a few years, there's been a lot of discussion about "bots" online. Over time, however, it's become a loaded and often misunderstood term. In this blog, you can read more from Yoel Roth, Head of Site Integrity, and Nick Pickles, Director of Public Policy Strategy, about bots, platform manipulation, what's prohibited by Twitter, and how the term "bots" is often misused.

### Misleading Information

Twitter's policy on misleading information has evolved to meet new challenges:

- We have a civic integrity policy that prohibits the use of Twitter for the purpose of manipulating or interfering in elections or other civic processes.
- In light of the unique public health threat posed by COVID-19, Twitter prohibits content that goes directly against guidance from authoritative sources of global and local public health information and which is likely to lead to offline harm. You can find more detail and criteria here.
- Finally, we prohibit the deceptive sharing of synthetic or manipulated media that are likely to cause harm – see blog.

We have begun to label Tweets that contain misleading information that falls within these three categories. These labels often lead to informational 'Moments' – annotated collections of Tweets that offer additional context.

### Public Interest Notices

We believe it's important people can read and speak about what world leaders say, even if they violate our rules. If a Tweet from a world leader does violate the Twitter Rules but there is a clear public interest value to keeping the Tweet on the service, we may place it behind a notice that provides context about the violation and allows people to click through should they wish to see the content. See more here.