Week 7: Friday

Ethical Issues in Technology

Unit 5: Platforms

Roadmap

~~Monday~~

Personalized feeds

Filter bubbles and echo chambers

Recommendation algorithms

~~Wednesday~~

[Case study] YouTube Recommendations

****Project Midpoint Due****

Friday

Who is responsible for content moderation?

Platform election response

[Current events]

Friday

Content Moderation

Should social media platforms be held responsible for moderating **user-generated** content on their sites?

Section 230: "Good Samaritan" Provision

(c) Protection for "Good Samaritan" blocking and screening of offensive material

(1) Treatment of publisher or speaker
No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.

(2) Civil liability

No provider or user of an interactive computer service shall be held liable on account of—

(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).[1]

A Framework for Moderation

- Moderation framework by Ben Thompson of Stratechery
- Discusses the banning of 8chan by Cloudflare
- Level of responsibility for each company is determined by their position on "The Stack"
- Discusses the legality vs. responsibility of moderation
- We will discuss a follow-up to this article **next week**

Platforms: Top of the Stack

At the top of the stack are the service providers that people publish to directly; this includes Facebook, YouTube, Reddit, and other social networks. These platforms have absolute discretion in their moderation policies, and rightly so. First, because of Section 230, they can moderate anything they want. Second, none of these platforms have a monopoly on online expression; someone who is banned from Facebook can publish on Twitter, or set up their own website. Third, these platforms, particularly those with algorithmic timelines or recommendation engines, have an obligation to moderate more aggressively because they are not simply distributors but also amplifiers.

ISP: Bottom of the Stack

It makes sense to think about these positions of the stack very differently: the top of the stack is about broadcasting — reaching as many people as possible — and while you may have the right to say anything you want, there is no right to be heard. Internet service providers, though, are about access — having the opportunity to speak or hear in the first place. In other words, the further down the stack, the more legality should be the sole criteria for moderation; the further up, the more discretion and even responsibility there should be for content:
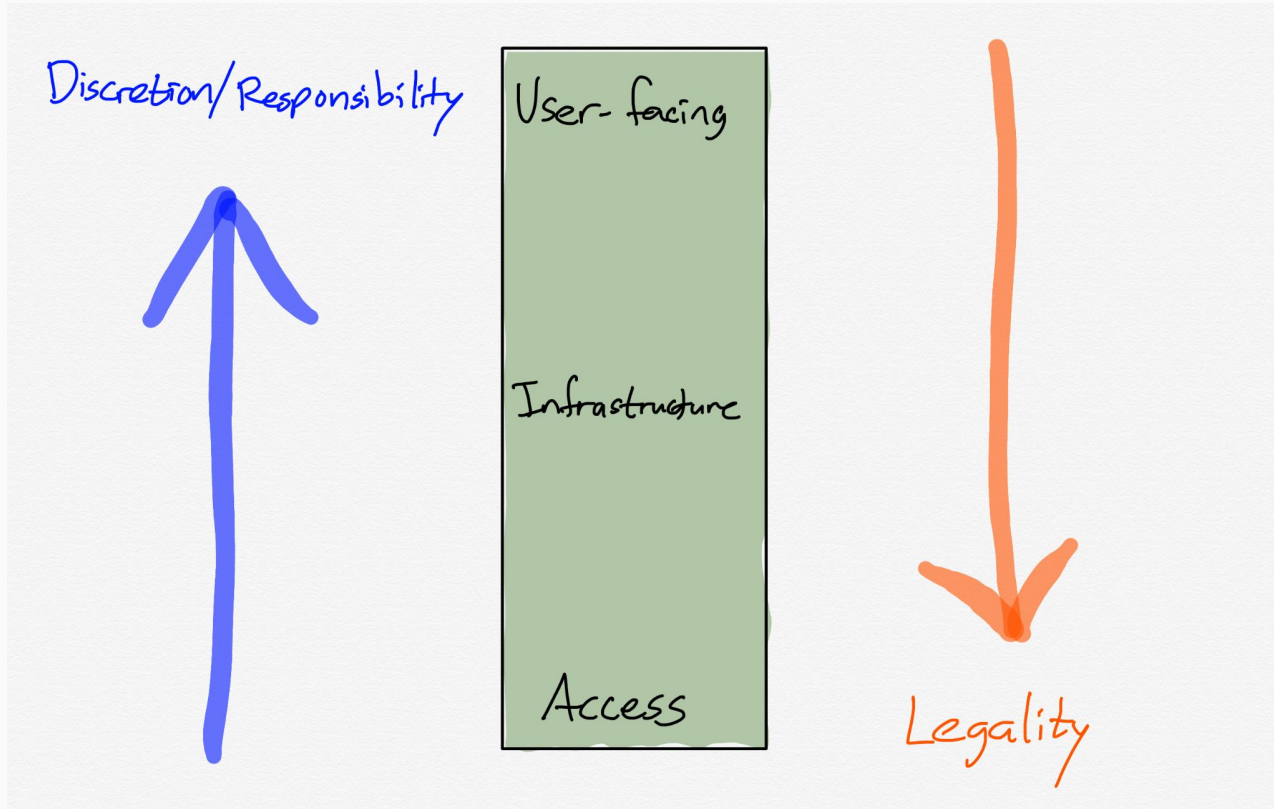
Breakout Rooms

- Meet with your discussion groups for ~5 minutes

- Discuss "A Framework for Moderation"

- Do you agree with the conclusions that Thompson makes regarding the responsibilities of companies at different levels?

- Could an ISP denying a customer the right to post online be considered a free speech violation?

- Are users able to freely move between platforms?

2020 Election Response: Facebook

A Look at Facebook and US 2020 Elections:

- preventing election interference
- creating rapid response tools and teams
- righting misinformation and voter interference
- addressing movements tied to violence
- adapting to the unique circumstances of this election
- improving transparency and privacy
- empowering voters through reliable information

https://about.fb.com/wp-content/uploads/2020/12/US-2020-Elections-Report.pdf

September 3, 2020

Today, Mark Zuckerberg announced additional steps we're taking to help secure the integrity of the US elections by encouraging voting, connecting people to authoritative information, and reducing the risks of post-election confusion.

- We won't accept new political ads in the week before the election.

- We'll remove posts that claim that people will get COVID-19 if they take part in voting, and we'll attach a link to authoritative information about the coronavirus to posts that might use COVID-19 to discourage voting.

- We will attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud.

- If any candidate or campaign tries to declare victory before the final results are in, we'll add a label to their posts directing people to the official results from Reuters and the National Election Pool.
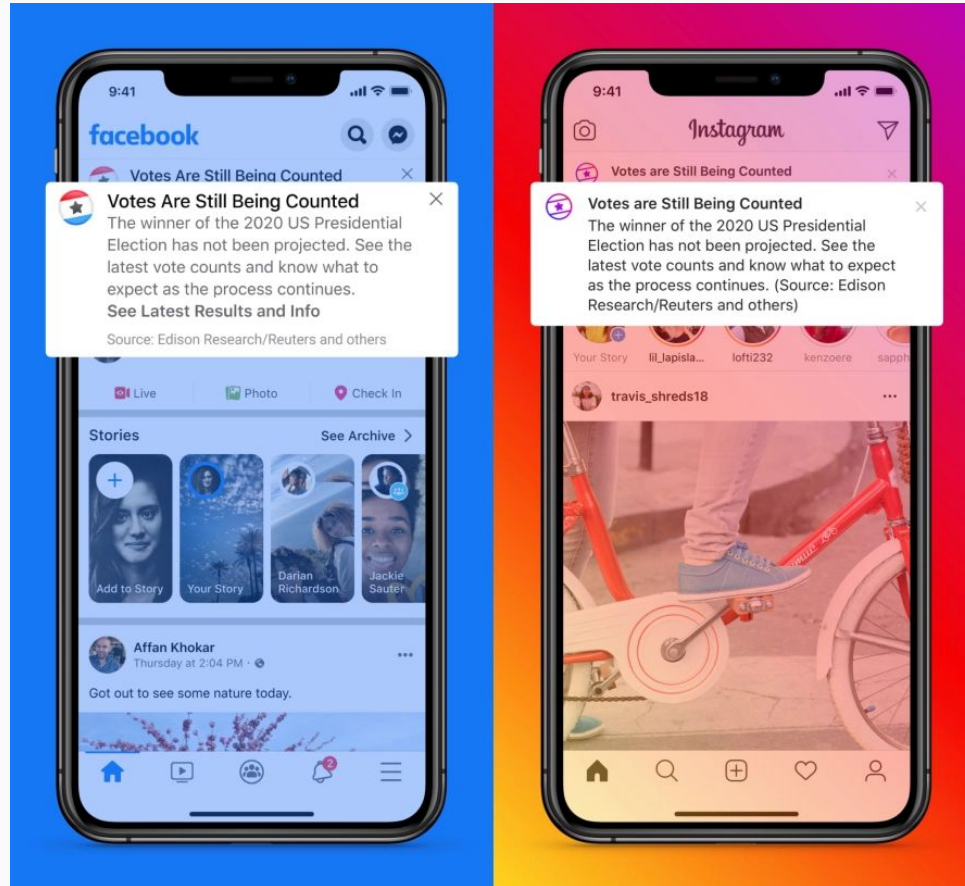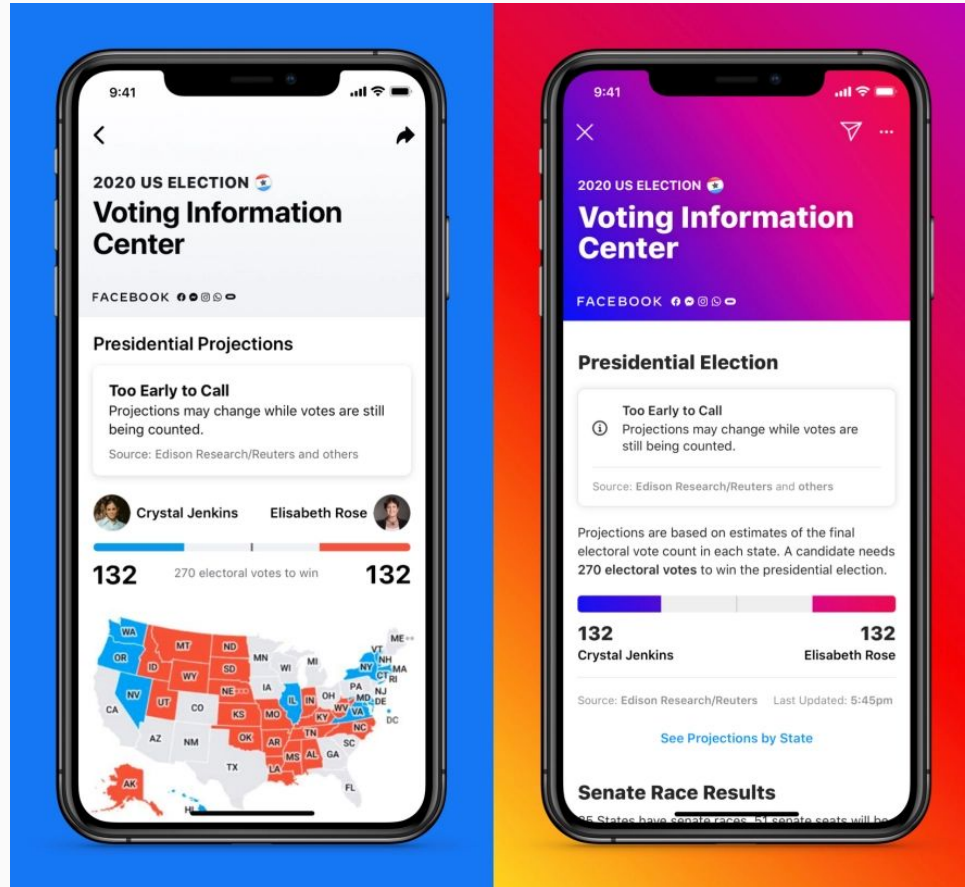
TECH • SOCIAL MEDIA

Facebook Acted Too Late to Tackle Misinformation on 2020 Election, Report Finds

Facebook could have prevented billions of views on pages that shared misinformation related to the 2020 U.S. election, according to a new report released Tuesday, which slams the platform for "creating the conditions that swept America down the dark path from election to insurrection."

The report, by the online advocacy group Avaaz, found that if Facebook had not waited until October to tweak its algorithms to stem false and toxic content amplified on the platform, the company could have prevented an estimated 10.1 billion views on the 100 most prominent pages that repeatedly shared misinformation on the platform ahead of the election.

2020 Election Response: Twitter

Actions taken by Twitter to prepare for 2020 election:

- Expansion of Civic Integrity Policy to handle misleading information
  - label false claims of candiate victory
  - remove tweets that encourage violence / election interference
- New prompts and more warnings on misleading Tweets
- Encourage quote tweets instead of retweets
- Change recommendations and trending topics for "added context"

Twitter: UX Design

Twitter: Misinformation label policy
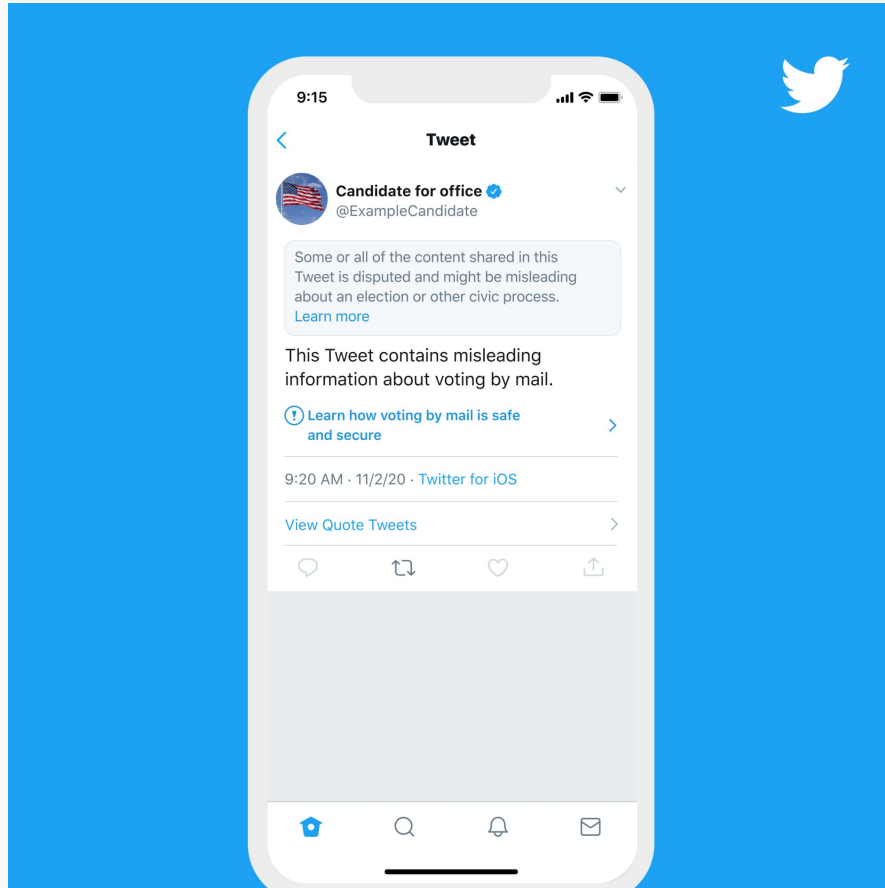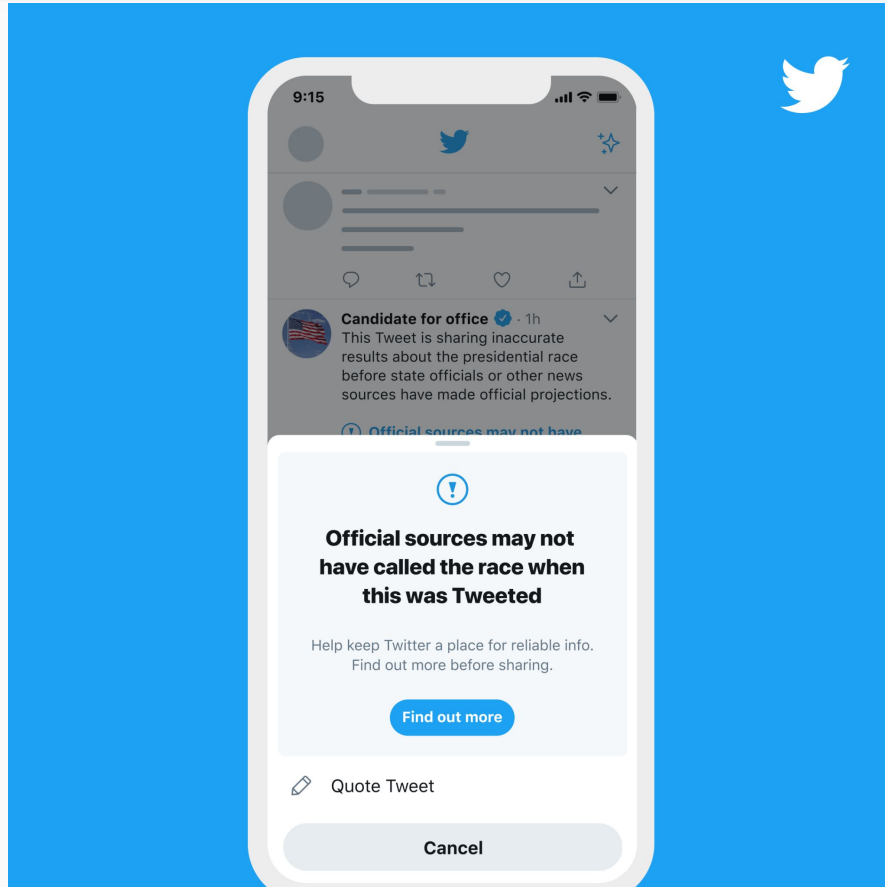
New prompts and more warnings on misleading Tweets

We currently may label Tweets that violate our policies against misleading information about civic integrity, COVID-19, and synthetic and manipulated media. Starting next week, when people attempt to Retweet one of these Tweets with a misleading information label, they will see a prompt pointing them to credible information about the topic before they are able to amplify it.

Tweets with labels are already de-amplified through our own recommendation systems and these new prompts will give individuals more context on labeled Tweets so they can make more informed decisions on whether or not they want to amplify them to their followers.

How will we identify these Tweets?

Our teams are using and improving on internal systems to proactively monitor content related to COVID-19. These systems help ensure we're not amplifying Tweets with these warnings or labels and detecting the high-visibility content quickly. Additionally, we'll continue to rely on trusted partners to identify content that is likely to result in offline harm. Given the dynamic situation, we will prioritize review and labeling of content that could lead to increased exposure or transmission.

Interactive Demo

Read through Twitter thread describing "fact-checking" process and addressing label backlash

https://twitter.com/TwitterSafety/status/1267986499112439809?s=20

Breakout Rooms

- Meet with your discussion groups for ~10 minutes

- Discuss the pros and cons of Twitter's moderation approach

- Have you noticed any effect of labeling on your own media consumption?

- Is it reasonable for Twitter to determine which sources to "amplify" for election and COVID information?

- Should trending topics and search results be moderated by Twitter?

2020 Election Response: YouTube

YouTube's plan for the 2020 election included:

- Removing videos that violate Community Guidelines
  - enforcement of Presidential Election Integrity policy
- Raise up authoritative information sources
  - promoting authoritative news channels
- Reduce recommendations of borderline content and misinformation
- Reward content creators with advertiser-friendly videos

Bloomberg

Screentime

YouTube Election Loophole Lets Some False Trump-Win Videos Spread

The site says election misinformation policies are 'generally working as intended.'
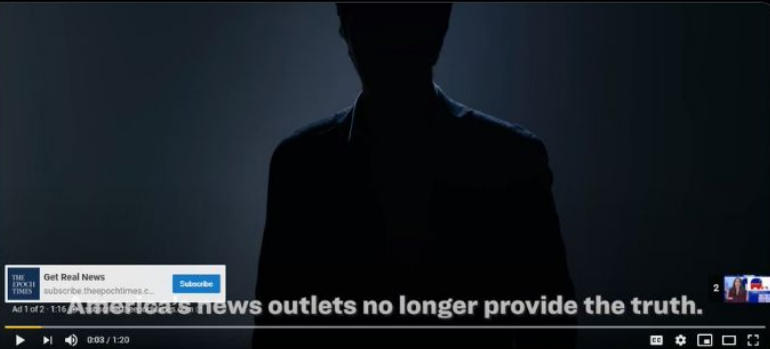
By Mark Bergen
November 10, 2020, 6:00 AM EST

YouTube: Media criticism response

Breakout Rooms

- [The 2020 U.S. elections on YouTube: Retrospective](#)

- [Election Fraud Narratives on YouTube](#)

- Meet with your discussion groups for ~10 minutes

- Review both of the above studies on YouTube's election response

- Why do the findings in these reports vary so dramatically?

- Was YouTube's 2020 election response adequate?

Closing Thoughts
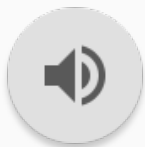
ADRIAN CHEN BUSINESS 10.23.2014 06:30 AM

The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed

Inside the soul-crushing world of content moderation, where low-wage laborers soak up the worst of humanity, and keep it off your Facebook feed.

[Audio Link](#)

- Is it ethical to have human content moderators?

- Is it ethical to have AI content moderators?

Current Events Talks

- Keep it brief! (2-3 minutes)

- Keep it relevant - relate to themes of the course

- Post follow-up in Discussion post (~500 words)

- Link to interesting / engaging / free (!) sources

Next Week - Platforms continued!

Monday

Privacy

Facebook-Cambridge Analytical data scandal

Data protection policies

Wednesday

[Case study] Trump Twitter Ban

World leaders and social media

Friday

Free speech on the internet

Moderation in infrastructure guidelines

[Current events]