



Political Influences on Content Policy

FRIDAY, DECEMBER 11, 2020 · READING TIME: 12 MINUTES

(I think almost all these examples, and more, have been written about in the media, but I thought it's worth documenting in a single note.)

TLDR:

1. Facebook's decision-making on content policy is routi

Chats

REDACTED FOR CONGRESS

1. **Facebook's decision-making on content policy is routinely influenced by political considerations.**
2. **The Communications and Public Policy teams are routinely asked for input on decisions regarding** (a) enforcing existing content policy, (b) drafting new policy, and (c) designing algorithms. Those teams often block changes when they see that they could harm powerful political actors.
3. **The influence of the Public Policy team is a common topic of complaint inside the Integrity org,** and Directors working in integrity have publicly stated they think our processes be changed to exclude them.
4. **We can and should set up a firewall between content-policy and other parts of the company.** A more concrete argument for this here.

Political Influence on Content Policy at Facebook

1. **Facebook has a huge team of people working on content policy who do great work.** Our moderation infrastructure is a massive achievement, and none of this is meant to cast doubt on that.
2. **Facebook routinely makes exceptions for powerful actors when enforcing content policy.** The standard protocol for enforcement and policy involves consulting Public Policy on any significant changes, and their input regularly protects powerful constituencies. Detailed examples are given below. Briefly:

1. Misinformation repeat-offender escalations seem to have regularly been influenced by input from Public Policy, exempting publishers on the grounds that they are “sensitive” or likely to retaliate. In the US it appears that interventions have been almost exclusively on behalf of conservative publishers.
 2. We have made one-off carve-outs in misinformation enforcement, apparently due to political pressure.
 3. In India a politician who regularly posted hate speech was exempted by the Indian Public Policy team from normal punishment explicitly for political considerations. A Facebook spokesperson would only go as far as saying that this “wasn’t the sole factor.”
- 3. Facebook routinely makes decisions about algorithms based on input from Public Policy.**
1. When significant changes are made to our algorithms (ranking, recommendations) they are usually reviewed by staff from Public Policy. Public Policy typically are interested in the impact on politicians and political media, and they commonly veto launches which have significant negative impacts on politically sensitive actors.
- 4. Final calls about content policy are routinely made by senior executives.** In multiple cases the final judgment about whether a prominent post violates a certain written policy are made by senior executives, sometimes Mark Zuckerberg. If our decisions are intended to be an application of a written policy then it’s unclear why executives would be consulted. If instead there was an unwritten aspect to our policies, namely to protect sensitive constituencies, then it’s natural that we would like executives to have final decision-making power.

5. **Our enforcement could be more independent and transparent.** I have not seen any compelling reason why Comms and Public Policy could not be removed from content policy processes.

Questions

Q: Is there a benign explanation for the involvement of Policy in decision-making?

Explanations of Public Policy's role are typically along these lines:

"Public Policy teams are important to the escalations process in that they provide input on a range of issues, including translation, socio-political context, and regulatory risks of different enforcement options." (<https://fburl.com/p25t51fl>)

It seems to me that the "translation" and "context" parts could easily be provided by other teams. The "regulatory risks" could refer either to complying with local regulations on speech, in which case an independent team could provide this service, or to concern about regulatory backlash towards Facebook.

Q: Why does Facebook allow Public Policy have influence over enforcement?

I think the explanation is relatively straight-forward: Facebook is in continuous negotiation with national governments over various issues, those negotiations are handled by the Public Policy team. Governments are also aware that Facebook has signif

I think the explanation is relatively straight-forward. Facebook is in continuous negotiation with national governments over various issues, those negotiations are handled by the Public Policy team. Governments are also aware that Facebook has significant influence over political discourse. I imagine sure that politicians routinely ask for special treatment, implicitly or explicitly, and as currently set up the Public Policy team has power to grant these requests because they have control over content policy decision-making.

Q: Do other media separate political and editorial judgment?

In major democracies most media organizations have some kind of institutional separation between editorial decision-making and the business side (“editorial independence”). I’m sure the separation isn’t perfect but it’s a far stronger structure than what we have at Facebook.

In less robust democracies I believe it’s common for the media to have a cooperative and transactional relationship with the government. I think that Facebook should hold itself to a much higher standard.

Q: Is Public Policy input necessary for obeying local laws?

There are many cases of tech companies obeying local laws: e.g. Germany’s hate-speech laws, US’s FOSTA-SESTA, Europe’s “right to be forgotten,” and Chinese censorship laws. Facebook clearly needs some some specialized group to comply with these laws, but this appears to be only a small subset of what Public Policy currently does, and I believe legal advice could be given independently from Public Policy org.

Q: Would the oversight board change things?

Q: Would the oversight board change things?

Facebook has recently set up an external "oversight board" which is intended to make binding decisions about individual pieces of content. However I don't believe it would have had authority to influence most of the cases discussed above: i.e. it will have no influence over ranking changes, over "repeat offender" designation, over "opinion" designation, or over "dangerous actor" designation. Thus I think that, although the effect is definitely positive, it is quantitatively small.

Q: How could arms-length enforcement work?

For enforcement of existing policies our decisions could simply be made with no input from Public Policy, Comms, or other external-facing teams. Public Policy can still maintain relationships, and explain the reasoning behind decisions, but would have no input into the decisions themselves. This is already the case in ads policy enforcement, as far as I can tell, where there is a clear separation of enforcement from Sales teams.

For development of new policies and evaluation of algorithm changes we could again remove Public Policy from everyday decision-making. Instead we could have general principles which Policy and Comms could contribute to, e.g. principles regarding impact of algorithms on politicians, but the actual fine-grained decisions regarding changes in algorithms would be entirely independent.

Q: What would happen if we had arms-length enforcement?

References

Product Changes

Significant launches in News Feed, News, and Recommendation all have XFN review which includes Public Policy. Those reviews often concentrate on the impact on politicians and major news sources. A common complaint is that explicitly political considerations enter these decisions.

Some examples of discussion:

2020-06: [REDACTED] (Engineering Lead for Civic Integrity):

“Separate content policy from public policy ... Today we force members of our policy org to simultaneously optimize for two goals. They have to arrive at clear, enforceable, and consistent choices in how they implement Mark’s principles around speech and the company’s values. However, they also have to anticipate and plan for the effects these choices will have on governments, our ability to run the company, and our regulatory environment.

2020-06: [REDACTED] (PM director, Civic Integrity):

“Separate Content Policy & Public Policy ... I have heard from many colleagues on the content policy team that they feel pressure to ensure their rec

“Separate Content Policy & Public Policy ... I have heard from many colleagues on the content policy team that they feel pressure to ensure their recommendations align with the interests of policymakers. They attribute this to the organizational incentives of having the content policy and public policy teams share a common root. As long as this is the case, we will be prematurely prioritizing regulatory interests over community protection moving swiftly to elevate our incredibly capable content policy team and making them an independent org is one of the most important reforms we can consider.

2020-05: [REDACTED] (ex-Chief Security Officer of Facebook)

“A core problem at Facebook is that one policy org is responsible for both the rules of the platform and keeping governments happy. ... It is very hard to make product decisions based upon abstract principles when you are also measured on your ability to keep innately political actors from regulating/investigating/prosecuting the company, especially when they have very concrete demands. ... Outside of the US, this is reflected in the fact that the local policy heads are generally pulled from the ruling political party and are rarely drawn from disadvantaged ethnic groups, religious creeds or castes. This naturally bends decision-making towards the powerful. ... This tension would better be reflected if the org structure had different VPs representing these equities and, hopefully, the one pushing principled decision-making winning more than their share.

2020-10: [REDACTED] (researcher)

“In recent months, time and again I’ve seen promising interventions from integrity product teams ... be prematurely stifled or severely constrained by key decision makers—often based on fears of public and policy stakeholder responses. ... I’ve

2020-10:

(researcher)

“In recent months, time and again I’ve seen promising interventions from integrity product teams ... be prematurely stifled or severely constrained by key decision makers—often based on fears of public and policy stakeholder responses. ... I’ve seen already built & functioning safeguards being rolled back for the same reasons.

As one small example, in July 2020 a set of changes to group recommendations was vetoed, explicitly on the basis of potentially upsetting political actors:

“A noticeable drop in distribution for these producers (via traffic insights for recommendations) is likely to result in high-profile escalations that could include accusations of shadow-banning and/or FB bias against certain political entities during the US 2020 election cycle.”

Hate Speech

2020-08: India “dangerous individual” designation.

A WSJ article claims that the head of Indian public policy exempted a powerful Indian politician from being designated a “dangerous individual”, explicitly due to the possible political damage to Facebook.

Facebook’s official response seems to acknowledge that this was a Chats

REDACTED FOR CONGRESS

Facebook's official response seems to acknowledge that this was an influence on the decision:

"A Facebook spokesman, [REDACTED] acknowledged that Ms. Das had raised concerns about the political fallout that would result from designating Mr. Singh a dangerous individual, but said her opposition wasn't the sole factor in the company's decision to let Mr. Singh remain on the platform."

<https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346>

There has been extensive internal discussion of this case (<https://fburl.com/8ezem42z> ; <https://fburl.com/p25t51fl>), which says that the designation is still under discussion, and has been for six months. As far as I can tell there has been no explicit refutation of the claim that this decision was influenced by fear of political consequences for Facebook.

2020: Trump Hate Speech

There have been a number of cases where senior executives made the final call about whether posts by the US President violated certain rules. Whether or not the posts violated the rules it's unclear to me why executives should be making this decision, rather than content-policy experts, unless it is due to political considerations.

More details at [2020-06-03 / FB & Twitter Response to May 2020 Trump Posts](#)

Misinformation

Discretion in the misinformation process.

Although Facebook's Misinformation process is nominally independent, depending on third-party fact checkers (3PFC), there are still multiple ways in which Facebook retains discretion:

1. Facebook reserves the right to judge whether content is classified as "opinion," which can exempt it from fact-checking.
2. Facebook reserves the right to exempt actors from "misinformation repeat offender" designation.
3. One-off exemptions from the usual process (described below).
4. Facebook has "newsworthiness" and "politician" exemptions for the misinformation policy, and the definitions of these categories are at the discretion of Facebook.

There is further discussion of cases in a *FastCompany* article.

2020: Many "Repeat Offender" designations were overturned after escalation by Public Policy

In early 2020 multiple tasks were filed by Public Policy, escalating "misinformation repeat
Chats

2020: Many "Repeat Offender" designations were overturned after escalation by Public Policy

In early 2020 multiple tasks were filed by Public Policy, escalating "misinformation repeat offender" designations. The majority of these were for large conservative publishers, including Breitbart, Charlie Kirk, Diamond and Silk and PragerU, and none were for liberal publishers. Some escalations explicitly mentioned that fear of political backlash was a contributing factor:

"this partner is extremely sensitive and has not hesitated going public about their concerns around alleged conservative bias on Facebook"

<https://fburl.com/hbqep7kh>

2019: Overturned fact-check on "abortion is never medically necessary."

In 2019 Facebook applied a one-off exemption from fact-checking to a post about abortion, after extensive political lobbying. This exemption was not based on any announced policy, and the designation seems to still be in effect.

The precise timing seems to be as follows:

1. The "Live Action" page uploaded a video with the claim "abortion is never medically

1. The "Live Action" page uploaded a video with the claim "abortion is never medically necessary"
2. A fact checker ("Health Feedback") rated this as False, and we labelled it as false.
3. The page lobbied for the fact-check to be reconsidered, and there was some internal discussion.
4. Republican senators reached out to Facebook to complain.
5. Facebook lodged a complaint with the IFCN (fact-checking network), arguing that the fact-checkers were affiliated with advocacy groups.
6. Facebook reclassified the post as "Not Eligible for fact checking." This was a "spirit of the policy" decision, based on a new policy that was in preparation but not yet released. (discussion by Zuckerberg @ <https://fburl.com/landun7u>)
7. The new draft policy was reconsidered, and never formally released (as far as I understand).
8. The IFCN investigation upheld the original fact-check, though they also said the fact-check should've disclosed the authors' affiliation with advocacy groups (<https://fburl.com/wu8xhbx7>).
9. The page continues to repost videos with the same claim, "abortion is never medically necessary."

Coordinated Behavior

Timeline

Chat

About

More ▾

CC

Like · Reply · 4d



2

KT

I think I have read a few of your posts and though its sad to see you leave the company, I wonder if there is scope for an opinion that is completely in contrast with yours and potentially the majority of the company.

There is an aggressive movement from the left and the right that is pushing for more and more censorship and forcing companies like FB & Twitter to do "more". More in this case has been inline with: de-platforming people, removing content, curating distribution, etc.

After carefully understanding your viewpoint, I am yet to hear your considerations on freedom of speech and expression which FB and Twitter are routinely infringing upon. Now I understand these are private companies and not public utilities and they can do what they want, but that doesn't discount that there has been severe miss on FoS majorly because the company is now intervening and removing speech.

You claim on one hand that FB is intervening at the senior levels but on the other hand, you want FB to be more opinionated. This is backwards IMO. Further, you are considering third party fact checkers (people or algorithms) to be idealistic without any political motivation and arguing for more of this stuff to become mainstream in the company, further hurting freedom of speech and expression on the internet.

There are countless examples to state here on the left and the right. In a lot of these cases, people on the left and right were unable to simply "protest" as access was taken away.

Is it viable to have a discussion that FB is actually becoming editorial day by day and censoring more & more instead of a platform where people can communicate whatever they want (minus direct threats of violence)?

Like · Reply · 4d · Edited



12

REDACTED FOR CONGRESS



appreciate this reply! agree this is a really hard tradeoff and I don't think my argument is fully coherent.

I think this is the crux, you say: "You claim on one hand that FB is intervening at the senior levels but on the other hand, you want FB to be more opinionated."

Yes, I claim both. I think that FB can be (1) opinionated; and (2) non-corrupt/non-partisan.

I guess you're skeptical that's possible to do both at once, given human nature? But you do think that non-partisanship is possible?



Like · Reply · 4d · Edited



here are examples of media institutions that i think are both relatively opinionated and relatively non-corrupt: WSJ, FT, the Guardian, the Economist.



Like · Reply · 4d · Edited

Was there an additional thought after "This makes me think our civic content could be significantly better if .
"?"



Like · Reply · 4d



oh geez when i push out notes fast the quality control deteriorates, i'm sorry, will fix.



Like · Reply · 4d



Why do you write like you're running out of time?... See More

REDACTED FOR CONGRESS

Timeline

Chat

About

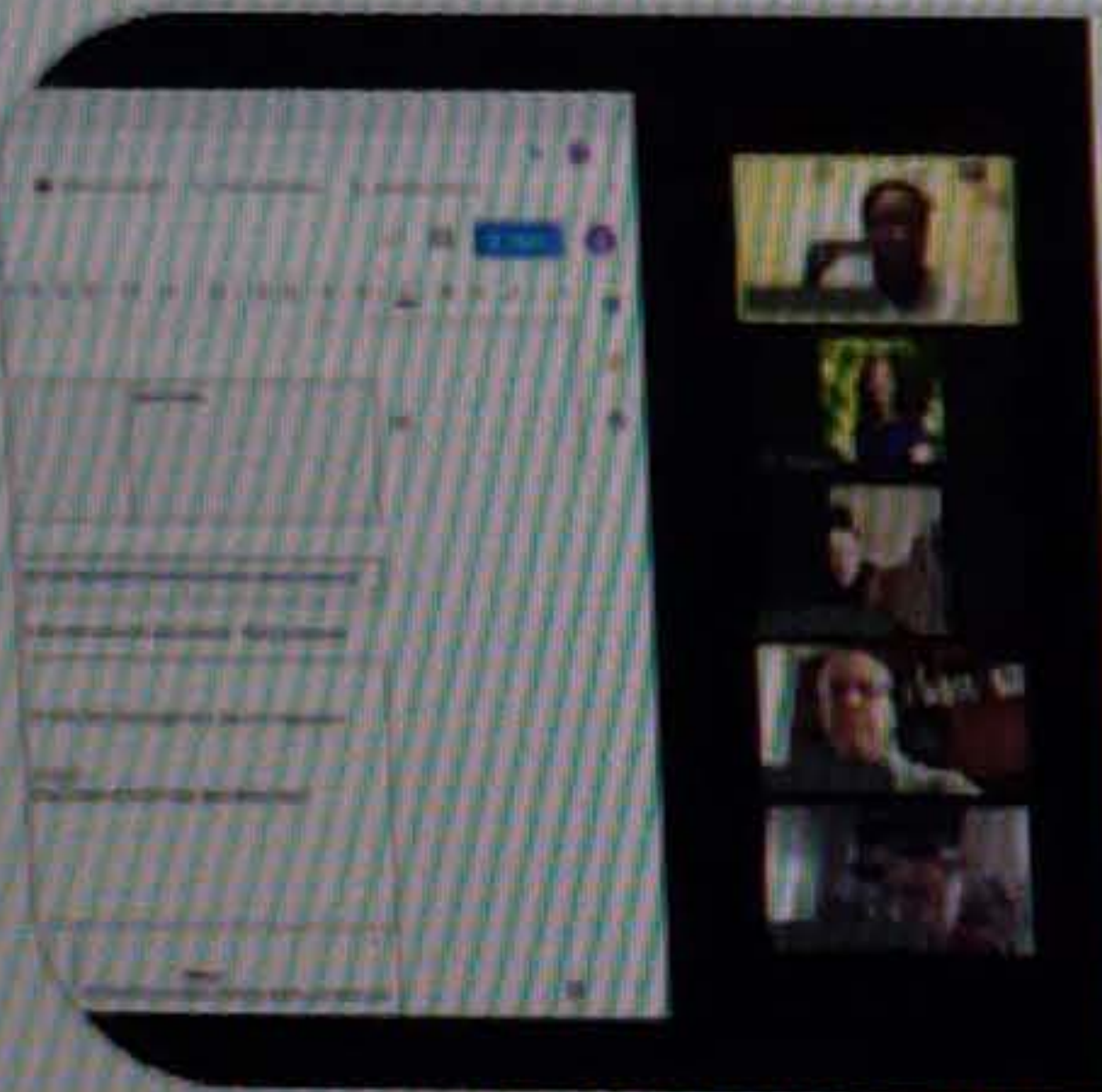
More

Note that Twitter *does* organizationally separate Public Policy from Content Policy functions (in terms of reporting chain). This is possible, we have simply chosen not to do it that way.

Like · Reply · 4d · Edited

29

The Stamos tweet thread appears to be gone now, but I found a copy of it at <https://twtext.com/article/1265394955515650050>



TWTEXT.COM

Thread by @wutrain, Sending to parents, guardians & students! Today starts the 1st full...

Like · Reply · 4d

2

thanks!

Like · Reply · 4d

Thank you for writing this up. I am so torn on this issue. On one hand, you make very important points and are probably right on all of them. On the other hand, as you point out, following these recommendations could hasten our own demise in a variety of ways, which might interfere all the other good we do in the world. How do you weigh these impacts?

Like · Reply · 4d

4

I've heard that argument a bunch of times and it always puzzles me – do we have any empirical or theoretical reason for thinking that simply trying to do the right thing is the wrong strategy longevity-wise?

...

12

REDACTED FOR CONGRESS

Timeline

Chat

About

More ▾

Thank you for writing this up. I am so torn on this issue. On one hand, you make very important points and are probably right on all of them. On the other hand, as you point out, following these recommendations could hasten our own demise in a variety of ways, which might interfere all the other good we do in the world. How do you weigh these impacts?

Like · Reply · 4d



I've heard that argument a bunch of times and it always puzzles me – do we have any empirical or theoretical reason for thinking that simply trying to do the right thing is the wrong strategy longevity-wise?

Like · Reply · 4d



I was just reacting to what [redacted] had written above - that this might make us more likely to be banned/broken up/ taxed.

I was also quite moved recently by reading the comments on Trump's 45 minute video on election fraud. So man... [See More](#)

Like · Reply · 4d



It turns out it actually does. When extremists migrate to platforms like Voat, Gab, Parler, etc. their reach diminishes significantly, and with it their ability to mobilize others for violence. Effectively the recruitment and popularization mechanics o... [See More](#)

Like · Reply · 1d · Edited



Is there data that only extremists are moving? I haven't seen that, can you share a link?

Like · Reply · 1d

REDACTED FOR CONGRESS

[Redacted] is there data that only extremists are moving?
I haven't seen that, can you share a link?

Like · Reply · 1d

Write a reply...



One of the things FB did when we created the Oversight Board was to define what its scope was. Now that it exists, we no longer have the power to alter their foundational documents unilaterally but - FB could enter into some kind of agreement w/ the Board whereby we delegate more powers to it - like some of the things you flagged: like ranking changes or any of the other 'shadow' community standards. I'm even in favor of FB delegating to an outside authority the ability to set content standards in the first place. With the Board, we sort of (emphasize on sort of) created a 'judicial' body to judge the quality of our enforcement. I think it stands to reason we also alienate our own "legislative" powers, as well. Ideally, this could go some ways to addressing the influence you're concerned with here. (not that, for the record, I agree completely with the examples you cite, but i do agree that our content policies should be fairly written and fairly enforced).

Like · Reply · 2d



Write a comment...



[Redacted] Statistics @ FB
December 11 at 2:18 PM · 📎

[Redacted] TRIPS Proxy Signals
December 11 at 2:12 PM · 📎

Note on Proxy/Surrogate Metrics for User-Level Surveys

REDACTED FOR CONGRESS

Tom Cunningham Integrity: Links, Research, and Discussion

December 11 at 1:19 PM

On De-Amplification

I had a discussion with [John Hegeman](#) earlier this week and I said I was skeptical about "de-amplification" in feed as an integrity tool. I promised him I'd try to write something down.

This is very partial, apologies if it comes across too strident. And apologies that I won't be around to admit mistakes in my analysis.

Deamplification, as I understand it, essentially penalizes the downstream terms in the value-model:

$$V = \gamma_1(\text{upstream}) + \gamma_2(\text{downstream})$$

Traditional integrity demotions just affect γ_1 , while deamplification affects γ_2 .

1. IT'S NOT AN EFFICIENT WAY OF ACHIEVING AGGREGATE GOALS.

This is probably obvious to most people, but I think it's worth stating clearly. Suppose the team-level goals are:

$$\text{engagement} - \beta \text{prevalence}$$

Where engagement could be sessions, MSI, etc. If those goals represent our true preferences, then deamplification is an inefficient way of achieving them. There's a long discussion about efficient ways that a value-model can maximize an objective function, with many nuances, but I think it's fair to say that if we wanted to achieve our aggregate goals then we wouldn't hard-code a downstream penalty.

More concretely: for a given level of engagement, **deamplification means higher prevalence**. We're paying a price, in terms of prevalence, for distinguishing upstream and downstream terms.

2. DOWNSTREAM INTEGRITY EFFECTS ARE ALREADY ACCOUNTED FOR IN THE VALUE MODEL.

The downstream term in the value model already reflects the integrity demotions that will be applied to my friends. If a piece of content has high $p(\text{misinfo})$, then $p(\text{downstream})$ will be correspondingly lower, due to our prediction models

REDACTED FOR CONGRESS

Timeline Chat About More ▾

2. DOWNSTREAM INTEGRITY EFFECTS ARE ALREADY ACCOUNTED FOR IN THE VALUE MODEL.

The downstream term in the value model already reflects the integrity demotions that will be applied to my friends. If a piece of content has high $p(\text{misinfo})$, then $p(\text{downstream})$ will be correspondingly lower, due to our prediction models anticipating the demotion.

This represents the theoretical best-case: I'm sure that in practice the recursiveness of demotions isn't perfect, but I don't think we know whether it causes too-much or too-little penalization of downstream terms.

If we thought there was inefficiency then we should tackle that by (1) studying it (e.g. through looking at whether residuals are mean-zero for misinfo), and (2) fixing the prediction models, not by putting a hard-coded penalty on downstream terms.

3. THERE'S AN INCENTIVE TO ROUTE AROUND DE-AMPLIFICATION.

I believe that experiments are still evaluated based on upstream+downstream engagement, and the downstream engagement effects are not discounted by integrity predictions. This means there's a discrepancy between the value-model and the way that experiments are evaluated.

If this is true then teams have an incentive to route around the de-amplification terms in the value-model.

E.g. suppose that de-amplification suppresses a certain category of clickbait content that contributes a lot to engagement. Then any feed engineer who can find a value-model tweak which correlates with that content will have a launch that will be approved.

This applies both to engineers and our automated tools (AE/AX) which look for features that can increase topline goals, and are unaware of the upstream/downstream distinction in the value model.

Thus we should expect the effect of de-amplification to get weaker over time.

4. WHY ARE WE DOING DEAMPLIFICATION?

I haven't followed the deamplification discussion closely at all, and I sincerely apologize if I'm mischaracterizing this.

REDACTED FOR CONGRESS

Timeline

Chat

About

More ▼

Thus we should expect the effect of de-amplification to get weaker over time.

4. WHY ARE WE DOING DEAMPLIFICATION?

I haven't followed the deamplification discussion closely at all, and I sincerely apologize if I'm mischaracterizing this.

What I have seen seems to emphasize these points:

1. Bad quality posts (e.g. misinfo) are disproportionately likely to have high downstream terms.
2. People shouldn't be shown bad-quality content just because *other* people want to see it.

Deamplification may help those problems, but it isn't an efficient way of addressing them:

1. If some subset of posts are more likely to be misinfo than others, we should confirm that the classifiers are well calibrated (remember that the condition is equal marginal quality, not equal avg quality), and if the classifiers are not calibrated we should fix that.
2. If we want to account for different reactions to quality we can fix that with a subjective term in our demotion (and we should make sure to incorporate subjective values it into our topline goals, not just our value-model, otherwise AE will route around it).

Perhaps there's some other problem which deamplification addresses, but apologies I don't know what it is.

6. INSTITUTIONAL REASONS FOR DOING INEFFICIENT DEMOTIONS.

I'm going very conjectural here, but I think there may be institutional reasons why we sometimes pursue apparently inefficient demotions.

I think this is true regarding the "bad for the world" demotion. It is an attempt to do two things: (1) reduce objectively bad content; (2) nominally satisfy subjective preferences. The conflict of these two demands caused it to be implemented in an incoherent way: [ref](#).

I would conjecture that there might be similar insitutional imperatives behind de-amplification: we want to increase demotion on misinformation (and other harmful stuff) but it's difficult to get sign-off on simply turning up the dial on these

Timeline

Chat

About

More ▾

I think this is true regarding the "bad for the world" demotion. It is an attempt to do two things: (1) reduce objectively bad content; (2) nominally satisfy subjective preferences. The conflict of these two demands caused it to be implemented in an incoherent way: [ref](#).

I would conjecture that there might be similar insitutional imperatives behind de-amplification: we want to increase demotion on misinformation (and other harmful stuff), but it's difficult to get sign-off on simply turning up the dial on these demotions.

So we instead we've started a series of projects which are nominally targetting user-value, and the "de-amplification" project fits that pattern. We then choose which of these projects to launch not based on whether they deliver user value (their nominal purpose), but instead primarily on whether they reduce prevalence of harmful content.

The result is an increased demotion of bad content, but at the cost of efficiency, conceptual clarity, and technical debt.

14

8 Comments 2 Shares Seen by 416

Like

Comment

Share

I disagree with your view of the goals and inconsistency of the "bad for the world" demotion (including the comments on the other post). You can read the eval criteria for this project [here](#), which gives a good intuition of what the team is trying to achieve (which is a personalized approach despite its name). Also, the response from Alison in the other post. This same team also took steps in 2019 to quantify and guardrail against **possible adverse effects**. While probably imperfect and with room to improve, both the metrics and approach are trying (and **achieving**) moves in sentiment metrics as measured by the TRIPS metrics.

Like · Reply · 4d

2

REDACTED FOR CONGRESS

[redacted] for the sake of clarity and so I can follow, when you say "including comments on the other post" what is the other post you're referring to?

Like · Reply · 4d

[redacted] see "ref" in OP

Like · Reply · 4d



[redacted] oh damn I completely missed that. Thanks!

Like · Reply · 4d

Write a reply...



[redacted] I think what is missed here is that in many cases we know our detection for misinfo-type integrity harms is awful. We don't have the option to just turn up demotions since we don't have good detection. We do know that reducing downstream virality for c... **See More**

Like · Reply · 4d · Edited



[redacted] agree that functionally one purpose of it is to reduce aggregate misinfo-type integrity harms, but I think we'd be better-off if we were explicit that that is the function, rather than framing it as a principled distinction between user-value and non-u... **See More**

Like · Reply · 4d



... REDACTED FOR CONGRESS

Like · Reply · 4d · Edited



agree that functionally one purpose of it is to reduce aggregate misinfo-type integrity harms, but I think we'd be better-off if we were explicit that that is the function, rather than framing it as a principled distinction between user-value and non-u... See More

Like · Reply · 4d



Thanks for putting this together Tom and as someone who has worked on a lot of deamplification, I can't disagree with a lot of what you said. At some level, it's - as you point out - somewhat a compromise to what is possible vs. what is ideal. I do think that integrity metrics are often leading indicators of user value - e.g. if you happen to hit misinformation by deprecating deep reshares, you are likely increasing the overall "truthfulness" of the ecosystem. And that shifts the overton window internally as to what is possible such that I don't think it's out of the question to simply start over on ranking civic content in a way that wasn't possible before deamplification work landed and showed impact. It's not the ideal or most direct path to getting to user value....but it's the best path we have had. Definitely appreciate these thoughts and please keep watching this space from afar. I'm hopeful we'll do what you suggest above more directly eventually.

Like · Reply · 4d



Write a comment...



REDACTED FOR CONGRESS