Vaccine Hesitancy Comment Demotion

Credits

- Author List: @Joo Ho Yeo, @Nick Gibian, @Hendrick Townley, @Amit Bahl, @Matt Gilles
- Thanks:

Executive Summary

- What's your goal?
  - Drastically reduce user exposure to vaccine hesitancy (VH) in comments
- What is the product change?
  - Utilize the existing v1 VH classifier (English) to demote comments on ranked comments, meaning that they are filtered from ‘most relevant’ but are still visible in other tabs (ex ‘most recent’)
- What are the benefits of this launch?
  - VPVs on Vaccine post English comments vh p80: -10.6±2.1%
    - Projected launch impact: -934.8K±194.4K vpvs
    - Authoritative vh p80 comment vpvs: -26.7 (±4.1)%
  - CEP on Vaccine post English comments vh p80: -11.1 (±1.8)%
    - Authoritative vh p80 comments CEP: -26.1±3.0%
  - Decrease in other engagement of VH comments including create, likes, reports, replies
    - Scuba grouped by VH
- What are the costs of this launch?
  - No significant cost is observed.
- Risks of this launch
  - Not all comments are actually vaccine hesitancy, but we'd aligned with Health Policy on this risk in the COVID Lockdown Decisions meeting 2 weeks ago - https://docs.google.com/presentation/d/1Qo3STGq75yf70-VkOAY61g0offfYaOB2o2df6Sury44/edit#slide=id.gca2fb195a7_11_0
- How could this be made more aggressive?
  - Use lower thresholds for interventions
- How could this be made more conservative?
  - Use higher Thresholds for interventions

Background

Experiment Launch Post

Comments are a major surface relevant to our B2V efforts. We estimate that the prevalence of VH comments in Authoritative Health Pages is 25.3% and for other pages 19.42%. Now that the v1 Vaccine Hesitancy classifier has been cleared for this usecase, reducing the visibility of these comments represents another significant opportunity for us to remove barriers to vaccination that users on the platform may potentially encounter.
We do not expect user controls to resolve this issue, although teams have recently shipped tools to give people the ability to choose who can comment. [Link](https://docs.google.com/presentation/d/1xIn5T1pg8kCyUJeEx_Ehevh4tnhQYMiwL1zszwll4/edit#slide=id.gbafe05fca7_10_6)

This is a Break The Glass lever that we'll replace with higher quality detection comes online. For now, we do not have other tools for remediating the high prevalence of vaccine hesitancy in Health comments

- More context from COVID Lockdown Decisions, where this was approved [Link](https://docs.google.com/presentation/d/1Q635TGq75yf76-VkOAY61gOdfYyaOB2o2dF6SUry44/edit#slide=id.gca2fb195a7_11_0)

**Classifier**

We utilized a text-only version of the [VH classifier that Amit Bahl developed](https://drive.google.com/file/d/1Q635TGq75yf76-VkOAY61gOdfYyaOB2o2dF6SUry44) to score English-language comments on vaccine posts. We then performed a hand-labeling exercise to evaluate the model output.

Use of classifier approved by pxfn (L1076303PRV) and bwc.

**Threshold Selection**

VPV-weighted precision and recall on thresholds (weighted reservoir sampling) and label [based on the latest policy approved VH labeling guidelines](https://docs.google.com/presentation/d/1xIn5T1pg8kCyUJeEx_Ehevh4tnhQYMiwL1zszwll4/edit#slide=id.gbafe05fca7_10_6).
VH Comments Text Classifier Performance

<table>
<thead>
<tr>
<th>cleaned_score_bucket</th>
<th>adjusted_precision</th>
<th>adjusted_recall</th>
<th>precision_lower</th>
<th>precision_upper</th>
<th>recall_lower</th>
<th>recall_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0 - 0.3</td>
<td>0.000000</td>
<td>0.085489</td>
<td>0.168845</td>
<td>0.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>0.3 - 0.5</td>
<td>0.0126605</td>
<td>0.084746</td>
<td>0.170880</td>
<td>0.054186</td>
<td>0.890621</td>
</tr>
<tr>
<td>2</td>
<td>0.5 - 0.6</td>
<td>0.197687</td>
<td>0.147376</td>
<td>0.261657</td>
<td>0.381893</td>
<td>0.739008</td>
</tr>
<tr>
<td>3</td>
<td>0.6 - 0.65</td>
<td>0.386031</td>
<td>0.305534</td>
<td>0.539776</td>
<td>0.264406</td>
<td>0.513907</td>
</tr>
<tr>
<td>4</td>
<td>0.65 - 0.7</td>
<td>0.462159</td>
<td>0.318081</td>
<td>0.530412</td>
<td>0.234323</td>
<td>0.462183</td>
</tr>
<tr>
<td>5</td>
<td>0.7 - 0.75</td>
<td>0.490524</td>
<td>0.200457</td>
<td>0.426854</td>
<td>0.143747</td>
<td>0.286414</td>
</tr>
<tr>
<td>6</td>
<td>0.75 - 0.8</td>
<td>0.588057</td>
<td>0.165619</td>
<td>0.514105</td>
<td>0.113932</td>
<td>0.233647</td>
</tr>
<tr>
<td>7</td>
<td>0.8 - 0.85</td>
<td>0.677055</td>
<td>0.102807</td>
<td>0.598812</td>
<td>0.256986</td>
<td>0.145819</td>
</tr>
<tr>
<td>8</td>
<td>0.85 - 0.9</td>
<td>0.730999</td>
<td>0.067448</td>
<td>0.648663</td>
<td>0.185614</td>
<td>0.046554</td>
</tr>
<tr>
<td>9</td>
<td>0.9 - 0.95</td>
<td>0.744750</td>
<td>0.029177</td>
<td>0.656591</td>
<td>0.028318</td>
<td>0.042050</td>
</tr>
<tr>
<td>10</td>
<td>0.95 - 1</td>
<td>0.795918</td>
<td>0.013449</td>
<td>0.680000</td>
<td>0.008156</td>
<td>0.018791</td>
</tr>
</tbody>
</table>

Notebook: https://www.internalfb.com/intern/anp/view/?id=495202
Post: https://fb.workplace.com/notes/3942483092467702
Classifier-based prevalence

- Of the non-null results, 168.775 K comments are under vaccine posts, which is 0.499% (33.625 M are not): https://fburl.com/scuba/ranked_comment_actions/ezd40b5u
- Since the vaccine hesitancy classifier only covers english comments size greater than 5, comments with a Vaccine hesitancy score is 48.355K. Of those 2.235 K are eligible for demotion: https://fburl.com/scuba/ranked_comment_actions/p2fcyml5

Among english vaccine post comments the classifier-based prevalence (with the p58 threshold 0.75):

- 4.3% of comment VPVs
- 4.3% of comment creates
- 6.4% of comment likes

(When adjusted for precision and recall, the prevalence is estimated to be 20~30% vpv)

Source: https://fburl.com/scuba/ranked_comment_actions/tlkb18sm

VH Comment Examples

<table>
<thead>
<tr>
<th>vaccine_hesitancy_score</th>
<th>comment_id</th>
<th>comment_text</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87189</td>
<td>261957062286503</td>
<td>So what are you spraying over north edmonton over residential homes, poison</td>
</tr>
</tbody>
</table>
| 0.8623                  | 193117605957898| Only if you want to travel.. won't be forced...but can't go anywhere unless you have it. 
So yes it will be forced |
| 0.9141                  | 1015857957229327| Better headline: “Vaccine thought to be 98% effective turns out to be 99.992% effective.” |
| 0.8663                  | 281743050286949| Exclusive: F ormer Pfizer VP to AFLDS: ‘Entirely possible this will be used for massive-scale depopulation’ |
| 0.95312                 | 1731887923657815| I’ve written to him and others. Either they dont read , dont research or just want to push the killer vaccines |
| 0.85873                 | 10159016877751830| @[100000552011150:2048:Kim Chrest] there’s a 99.2% that you won’t. That is far above the 95% effective rate that we were told in the beginning. |
| 0.85002                 | 10158210334681728| I’m in a group whose families have had the “Jab”. Doesn’t matter which you have, all are causing terrible side effects and many dying. |
| 0.85447                 | 304527971035699| The vaccine is not fda approved but rather approved for emergency use only. Legally employers and others can not make it mandatory. |
| 0.94851                 | 4498348683527934| In other words, the Namibia human rights advocate, is calling for suspension of the rights of health workers and pushing for forced vaccinations. |
| 0.85625                 | 293217805580344| didn’t work for the boarder worker....had both vaccines and still got covid again..... |
| 0.87042                 | 10158393963727183| @[100002882938982:2048:Eifz O’Hare] how many have died from covid? How many have died from vaccine? |
| 0.96451                 | 10159572429229015| @[796927924:2048:Dave Harvey] A year ago we didn’t know how successful any anti-COVID vaccine, let alone mRNA vaccines, would be in the field. Not really joking... Day b4 yesterday my uncle died few days after vaccination in Washington due to clotting .. aged 71 .. Diagnosed COVID19 |
| 0.89849                 | 1200292507059482| Not really joking... Day b4 yesterday my uncle died few days after vaccination in Washington due to clotting .. aged 71 .. Diagnosed COVID19 |
| 0.91939                 | 2943603365923180| @[100000411263327:2048:Heather Keedy Bateman] Too bad they aren’t reporting all the deaths and serious adverse effects to the public! |
| 0.87968                 | 10159566195034228| @[100000814050919:2048:Lisa Sanders] 71 out of 1.4 million fully vaccinated people in Missourri isn’t too bad in my opinion. .000005% |

Experiment

QE: https://www.internalfb.com/intern/qe2/cref_v5_universe/vh_v1_en_demotion_v2/setup/config

QE is set up with a control group and the following test groups. Demotion is done to read time score at level -1:

- demote comments with v1 vaccine hesitancy classifier score of 0.75 (p58) or above with demotion strength of 0.01
- demote comments with v1 vaccine hesitancy classifier score of 0.75 (p58) or above with demotion strength of 0.001
- demote comments with v1 vaccine hesitancy classifier score of 0.8 (p58) or above without demotion strength of 0.01
demote comments with v1 vaccine hesitancy classifier score of 0.8 (p≥0.8) or above with demotion strength of 0.001

Ranking Changes

https://www.internalfb.com/intern/configurator/edit/?path=comment_ranking%2Fcomment_ranking_configs%2Fvh_v1_en_demotion_p58_001.cconf

demote_vh_en =>

```
{
    "signals": {
        "vh_score": 871
    },
    "signal_defaults": {
        "vh_score": 0
    },
    "phase": 1,
    "signal_blender": "if(gte($vh_score, 0.75), 0.001, 1)",
    "signal_rules": []
}
```

experiment_read_time_score =>

```
{
    "signal_blender": "mul($read_time_score, $demote_vh_en)"
    "phase": -1,
    "signals": [],
    "signal_rules": ["read_time_score",
    "demote_vh_en"
    ],
    "signal_defaults": {
        "read_time_score": 1,
        "demote_vh_en": 1
    }
}
```

Demotion Example

https://www.facebook.com/CDC/photos/a.184668026025/10159076965856026

While use of Johnson & Johnson’s COVID-19 Vaccine remains paused, you can still be vaccinated with either the Pfizer BioNTech or Moderna vaccine.
Can still be vaccinated with either the Pfizer-BioNTech or Moderna vaccine.

Make plans to get a vaccine as soon as one is available to you. Keep your vaccination appointment if you have one. If you were scheduled to get the J&J vaccine, work with your vaccination provider to reschedule to get a different COVID-19 vaccine.

Jenni Borders Why the push for 2 doses when CDC itself published efficacy of 80% after 2 weeks with just one dose of Pfizer or Moderna? The flu shot is 40 to 60% effective every year and the flu actually hospitalizes and kills children and elderly every single year here in the United States.

Like · Reply · 3d

Anna Di Franco Why is the CDC pushing masks when there is scientific evidence showing they do not protect us but instead cause bodily and psychological harm

https://www.outkick.com/stanford-study-says-masks-are...

Like · Reply · 2d

Sharon Mahany One question you might ask... Why are social media groups that are reporting firsthand accounts of side effects of the covid vaccines being shut down? Do they not want people to make an informed decision?
April Marie No it's not! Stop telling lies!! You know you're in cahoots with big pharma Bill Gates and other big names... you're going to kill people off the Earth shut the hell up!

Like · Reply · 1d

Comment type: comment_on_source_post
Metric Tags: survey_p_high, objective_high, relevance_high, top_fan_badge_eligible, score_red, civic, ranked
Position Change: 15
Whitelisted Features:
comment_text_vh_v1_score(871): 0.66857151985168
842197446646352 (Entity Tool)
Click to Dive Deeper

Rolanda Shaniqua Wait... Getting a quickly created pharmaceutical vaccine that we know has some issues right now, that has absolutely no data on long term effects, is not fully FDA approved, and is exempt from any liability for injury or death, is safer than keeping yourself naturally healthy with a good diet, natural supplements, and exercise? If people want to get this vaccine, fine, but why is the CDC resorting to stretching the truth even more with statements like these?

Like · Reply · 2d · Edited

Comment type: comment_on_source_post
Metric Tags: survey_p_high, objective_high, relevance_high, p_report_weak, top_fan_badge_eligible, score_red, civic, ranked
Position Change: 4
Whitelisted Features:
comment_text_vh_v1_score(871): 0.86716830730438
10159076986751026 (Entity Tool)
Click to Dive Deeper

Amanda Kay Harvey-Armstrong I get my second dose on the first. I can't wait to be fully vaccinated for this. The more people vaccinate the better. Maybe if we get enough people vaccinated
things MIGHT turn back to somewhat normal.

Josee Alain I feel bad for those that got the J&J vaccine having been told to take the first vaccine available to them, now they are waiting to see what comes out of this vaccination pause...

Melissa Albert Yet in Canada they're giving the Astra Zeneca, and no Moderna or Pfizer available. And they're canceling appts due to shortages. 😞 And back into lockdown if you aren't an airport or corporation or government or rich.
Results

Analyzing the p58 and demotion strength 0.01 (though all results are similar)

Deltoid

Vaccine Heistancy Prevalence Metrics

https://fburl.com/deltoid3/754nhd0z

- CEP on Vaccine post English comments vh p80: 11.1 ±1.8%
  - Authoritative vh p80 comments CEP: 26.1±3.0%
  - Non-authoritative vh p80 comments CEP: 9.9±2.3%
• VPVs on Vaccine post English comments vh p80: **-10.6±2.1%**
  - Projected launch impact: -934.8K±194.4K vpvs
  - Authoritative vh p80 comment vpvs: -26.7 (±4.1)%
    - Projected launch impact: -402.4K±71.3K
  - Non-authoritative vh p80 comment vpvs: -9.0±2.7%
    - Projected launch impact: -526.5K±163.4K

Other Signals:

• No detectable difference in misinfo frx:
  - https://fburl.com/daiquery/a1yg0xhx

• No detectable difference in high or low quality comment VPVs:
  - https://fburl.com/scuba/ranked_comment_actions/itgnv6zs,
  - https://fburl.com/scuba/ranked_comment_actions/r5z2koh3

• No detectable difference in FRX overall of any type:
  - https://fburl.com/deltoid3/1h64b6oi

Overall Engagement
https://fburl.com/deltoid3/iitnq125

Grouped by Reply
https://fburl.com/deltoid3/o4hnqgsd

Grouped by Reply time series
https://fburl.com/deltoid3/h9snr3uy
### Grouped by High/Low Relevance

[Link](https://fburl.com/deltoid3/xxi2mu5c)

### Grouped by Is bullying

[Link](https://fburl.com/deltoid3/jlyfs4c5)

### Grouped by is hate speech

[Link](https://fburl.com/deltoid3/vaa6rzzf)
**Scuba grouped by VH**

https://fburl.com/scuba/ranked_comment_actions/5zbwxf9

-75.6% create
-42.5% like
-34.2% report
-32.6% vpv
-88.6% replies

**Next Steps**

- Team holdout
- Live videos? Other surfaces?

**Follow Up Question Answers:**

1. What is the percentage of vaccine
   a. Of the non-null results, 168.775 K comments are under vaccine posts, which is 0.499% (33.625 M are not): https://fburl.com/scuba/ranked_comment_actions/ezd40b5u
      i. This aligns with the vaccine post vpvs / all post vpvs which is ~0.5%
         1. https://fburl.com/daiquery/grmcvdqx
   b. Since the vaccine hesitancy classifier only covers english comments size greater than 5, comments with a Vaccine hesitancy score is 48.355K. Of those 2.235 K are eligible for demotion: https://fburl.com/scuba/ranked_comment_actions/p2fcyml5
i. So the final eligible comments for demotion are 0.00661% of ranked comments

2. Change in "likes"
   a. For the vaccine hesitant comments we are demoting, we are reducing -2.64K likes (-42.5%), which would be 176K likes since the test size is 1.5%
      i. https://fburl.com/scuba/ranked_comment_actions/oldlnwjv
   b. Grouped by is_repy, time series
      i. https://fburl.com/deltoid3/xturut3g
   c. Grouped by survey_high, survey_low
      i. https://fburl.com/deltoid3/3a2ijqyw

* https://fb.workplace.com/groups/health.integrity/permalink/875731499946842
**https://fb.workplace.com/groups/health.integrity/permalink/505046790348650
***https://fb.workplace.com/groups/health.integrity/permalink/505046790348650