

Viral Dynamics, Algorithms, & the Information Environment

Jennifer Allen

NYU Stern School of Business



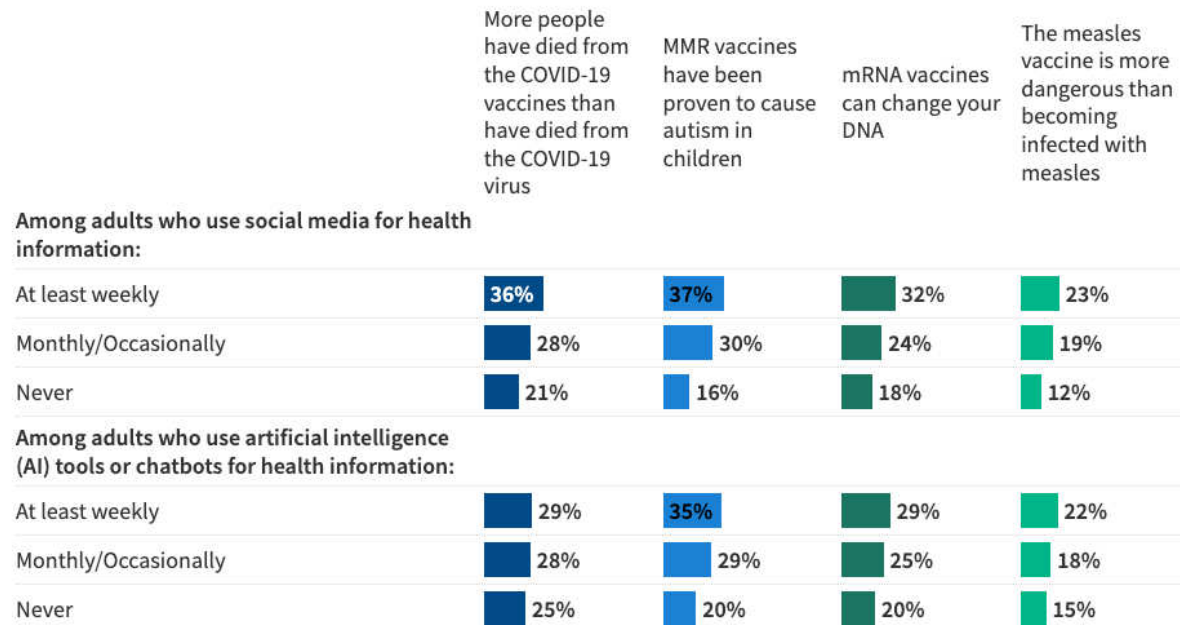
Learning Objectives

- Distinguish vaccine misinformation and vaccine skeptical content, and **why gray area skeptical content** is important
- Understand **three types of information spread** on social media: subscription, network, and algorithmic
- **Understand tools** to access social media data
- **Content moderation red-team** activity to demonstrate challenges of limiting vaccine hesitancy stemming from social media

Figure 5

Use of Social Media and AI Is Tied to Belief in Vaccine Myths

Percent who say the following false claims about vaccines are **definitely** or **probably** true:



Note: See topline for full question wording.

Source: KFF Tracking Poll on Health Information and Trust (May 7-31, 2026) • [Get the data](#) • [Download PNG](#)

KFF

What social media do you use?



What social media content has the most
impact on on vaccine hesitance?

Which of these would increase skepticism?



Which of these received the most views?



601 views

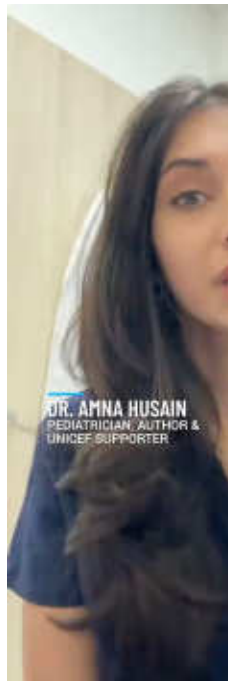


741 views



2.3 million views

Which of these would violate TikTok's policies?



601 views

Our rules on harmful misinformation

Our [Community Guidelines](#) prohibit harmful misinformation about health, elections, climate change and more. When content goes against these rules, we remove it or make it ineligible for the For You feed as outlined in our policies (learn about our policies [here](#)). We take action on misinformation regardless of intent. Our policies apply to both intentional "disinformation" as well as harmful misinformation that may not have been shared with the goal of deceiving people.

To help ensure our moderation decisions are grounded in facts, we work with independent [fact-checking partners](#) who assess whether information is true, false, or unverified. If you see misinformation that might break our rules, please [report](#) it.

When we take action on your content, we'll let you know through an inbox notification or within the "[Account Status](#)" page in your in-app Safety Center. You can also submit an [appeal](#) if you think the moderation decision was a mistake. When you report harmful misinformation, you can track the report's outcome in the "Report Records" page in your Safety Center.

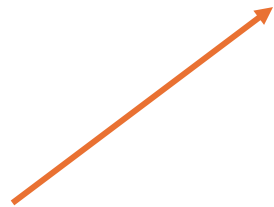
741 views



2.3 million views

Impact

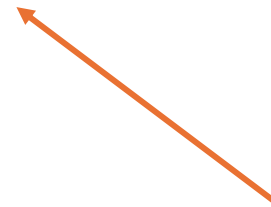
Does social media content impact vaccine hesitancy?
Which content has the biggest impact on vaccine hesitancy?



Reach

What kind of vaccine content do people see
on social media?

Who does it reach?



Persuasive Effect

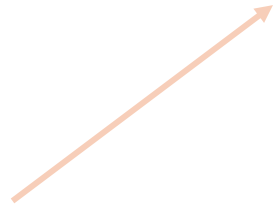
Which types of vaccine information influences
vaccine hesitancy?

What is the overall **impact** of vaccine-skeptical content on social media on vaccine hesitance?

1. Influence → what content induces vaccine hesitancy?
2. Reach → what determines is popular on social media?

Impact

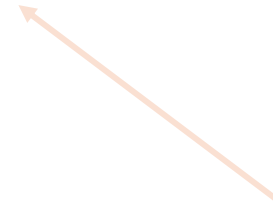
Does social media content impact vaccine hesitancy?
Which content has the biggest impact on vaccine hesitancy?



Reach

What kind of vaccine content do people see
on social media?

Who does it reach?



Persuasive Effect

Which types of vaccine information influences
vaccine hesitancy?

Influence: what types of content change
vaccine hesitancy?

Persuasion background

- In general, persuasive content on social media has **minimal effects** (Athey et al, 2023)
- Changing minds is very hard!
- Evidence that **vaccine misinformation** can decrease vaccine intentions, though mixed
 - Large effect (6pp compared to true)
 - Lomba et al (2021)
 - Null effect
 - De Saint Laurent et al (2022), Porter, Velez, and Wood (2022)
- Other drivers?

Harmful-to-Health Drives hesitancy

- Study which tested 130 pieces of vaccine related content on social media (Allen et al, 2024)
- **Harmful-to-health** predicts treatment effects
- Effect persists *even controlling for veracity*
- Asymmetric → little effect of positive content



Bill Filed in Washington Would Authorize 'Strike Force' To 'Involuntarily Detain' Unvaccinated Families

The Washington State Board of Health may soon amend state law to authorize the involuntary detainment of residents as young as 5 years old in Covid-19 "internment camps" for failing to comply with the state's experimental vaccine...

The Gateway Pundit - Alice Power - Jan 8, 2022

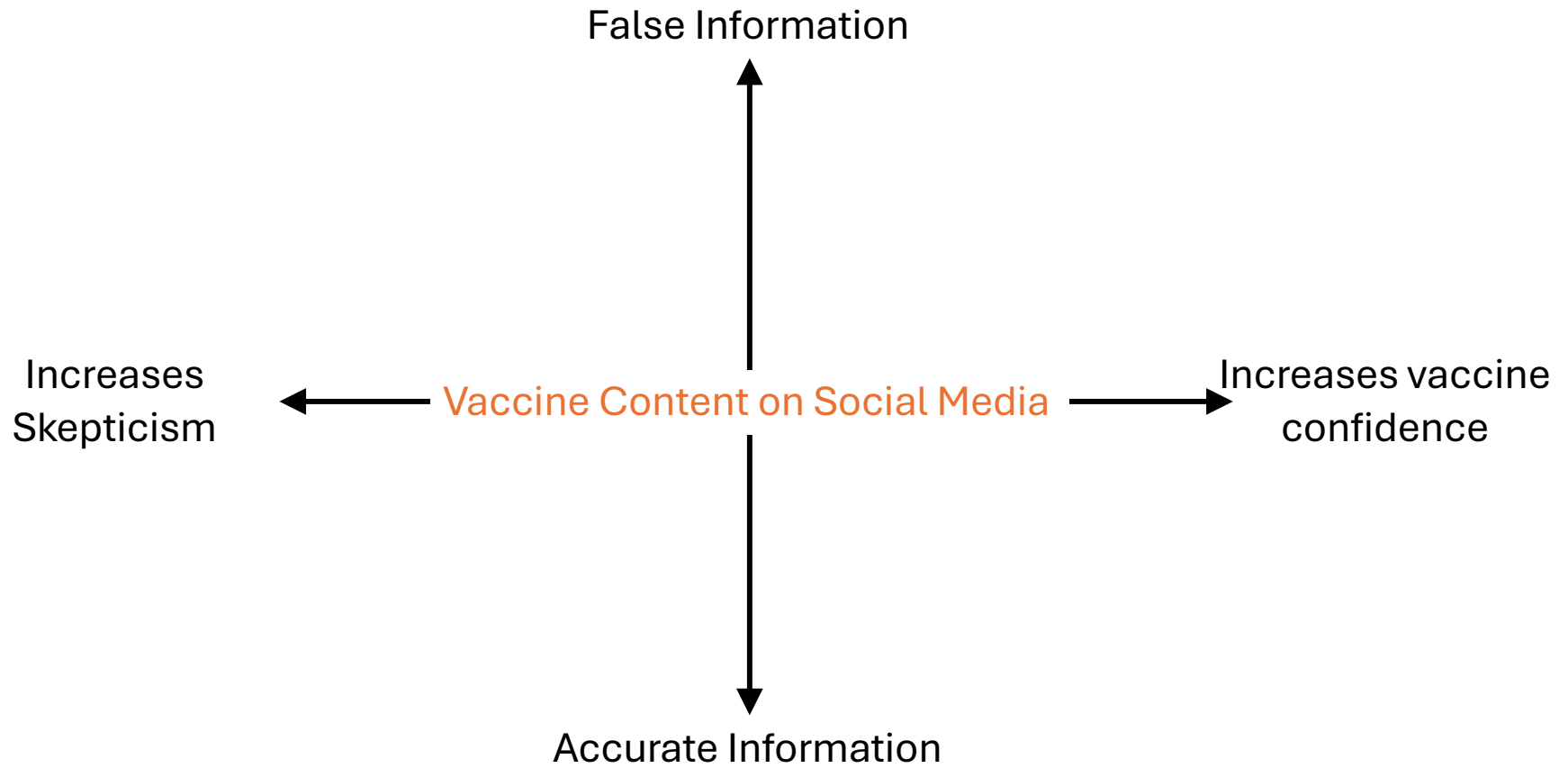
Suggests Vaccines are Helpful vs. Harmful to Health
(1 = Very Helpful, 5 = Very Harmful)

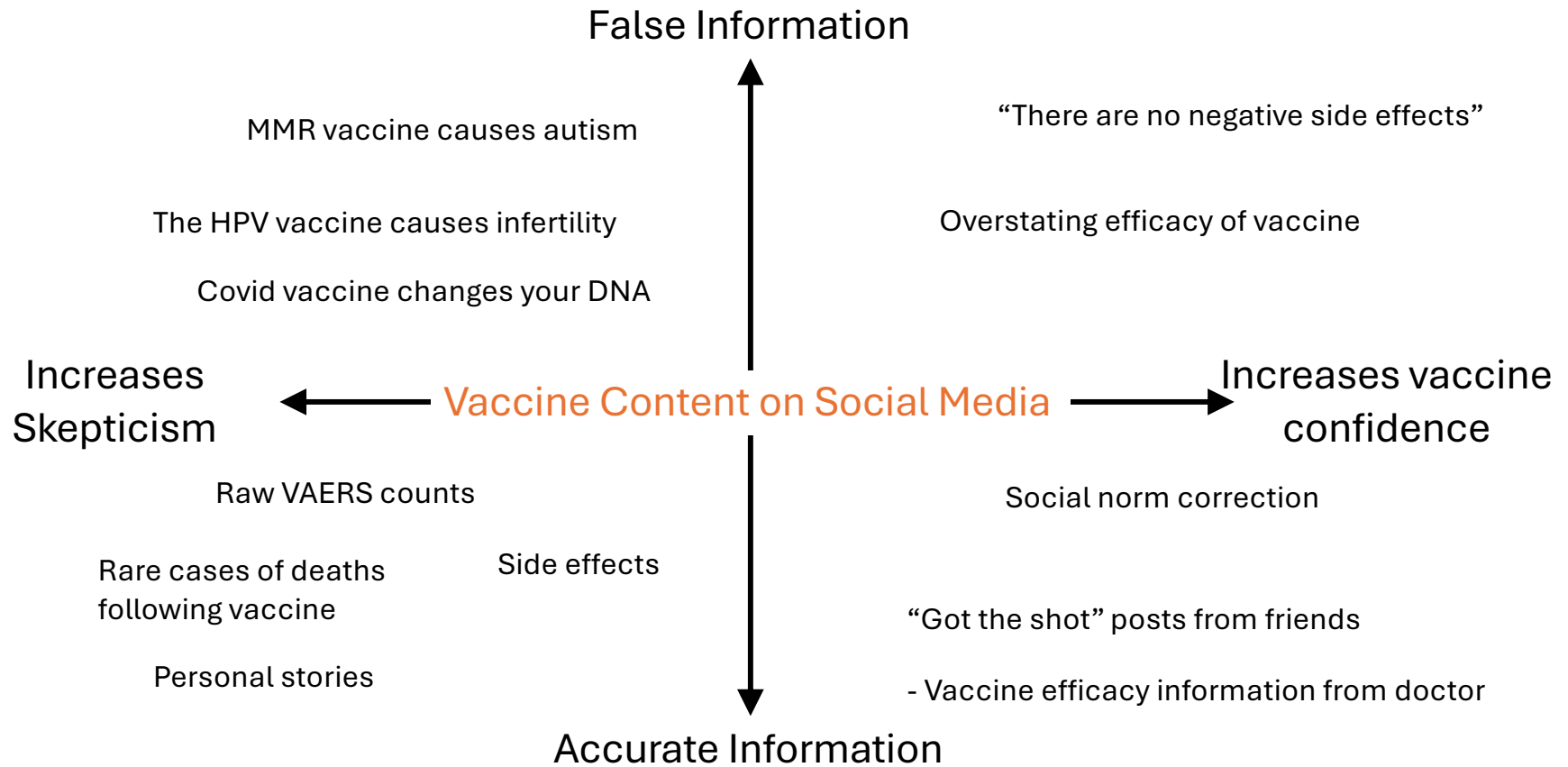
“Gray area” vaccine content



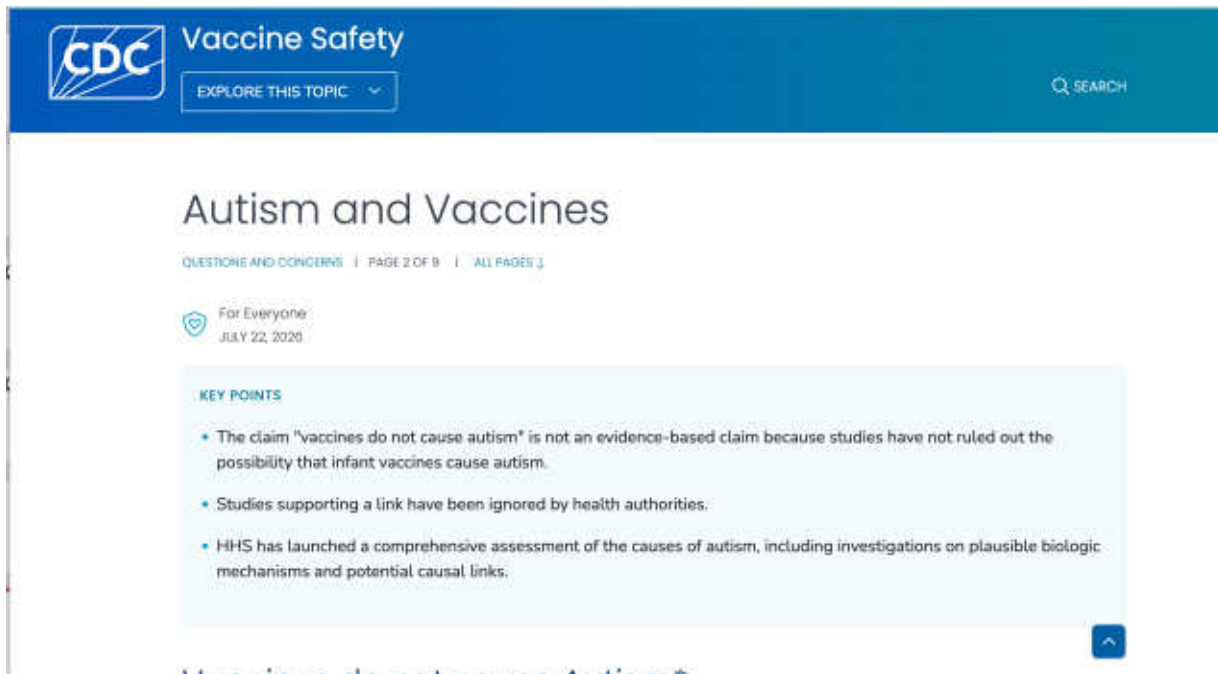
VS







Authoritative Sources are Shifting



The screenshot shows the top of a CDC website page. The header is blue with the CDC logo on the left, the text 'Vaccine Safety' in the center, and a search bar on the right. Below the header, the main title 'Autism and Vaccines' is displayed. Underneath the title, there are links for 'QUESTION AND CONCERNS', 'PAGE 2 OF 8', and 'ALL PAGES'. A shield icon with the text 'For Everyone' and the date 'JULY 22, 2020' is visible. A light blue box contains the heading 'KEY POINTS' followed by three bullet points. A small blue square with an upward arrow is at the bottom right of the key points box.

CDC Vaccine Safety

EXPLORE THIS TOPIC

SEARCH

Autism and Vaccines

[QUESTION AND CONCERNS](#) | [PAGE 2 OF 8](#) | [ALL PAGES](#)

 For Everyone
JULY 22, 2020

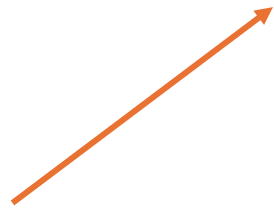
KEY POINTS

- The claim "vaccines do not cause autism" is not an evidence-based claim because studies have not ruled out the possibility that infant vaccines cause autism.
- Studies supporting a link have been ignored by health authorities.
- HHS has launched a comprehensive assessment of the causes of autism, including investigations on plausible biologic mechanisms and potential causal links.

Reach: How content spreads on social media

Impact

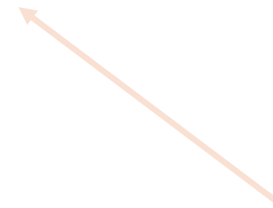
Does social media content impact vaccine hesitancy?
Which content has the biggest impact on vaccine hesitancy?



Reach

What kind of vaccine content do people see
on social media?

Who does it reach?

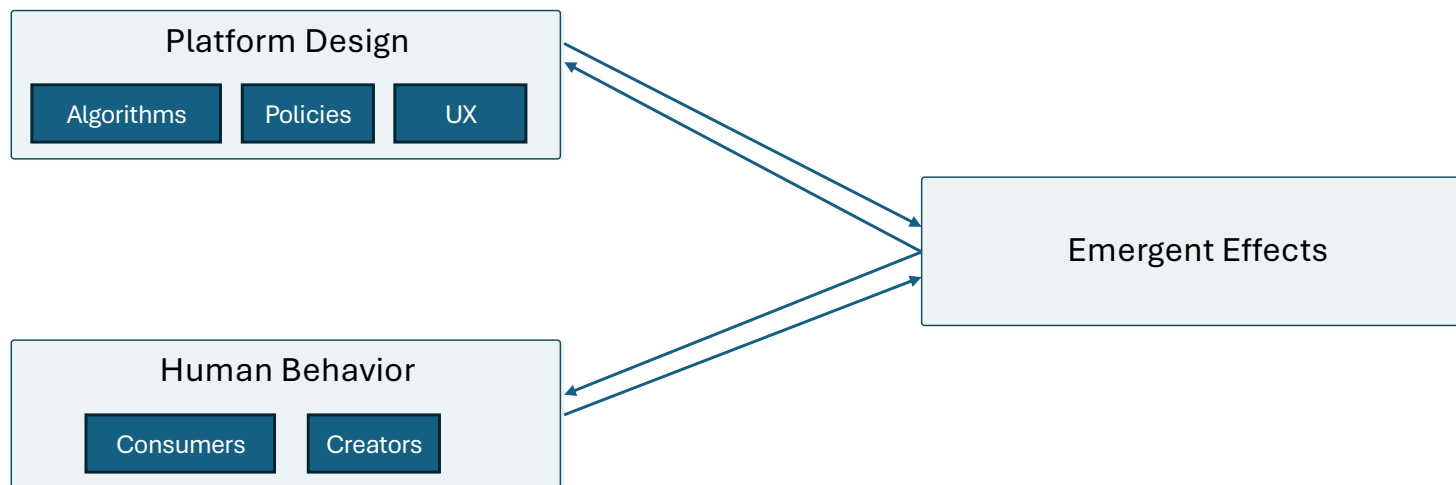


Persuasive Effect

Which types of vaccine information influences
vaccine hesitancy?

Question: How does social media decide what content to show you?

Social Media Is a “Complex System”



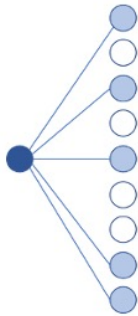
Social media pushes
vaccine skeptical
content



People are interested in
vaccine-skeptical content

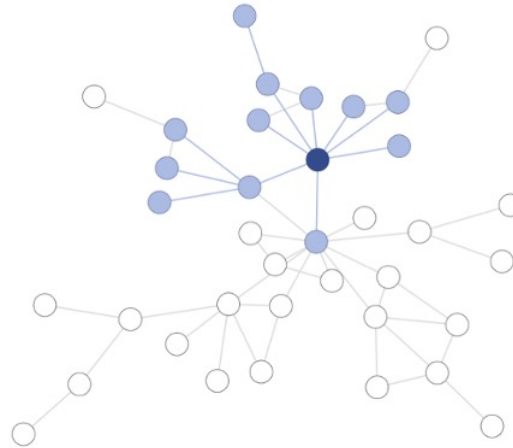
Types of Information Propagation

Subscription



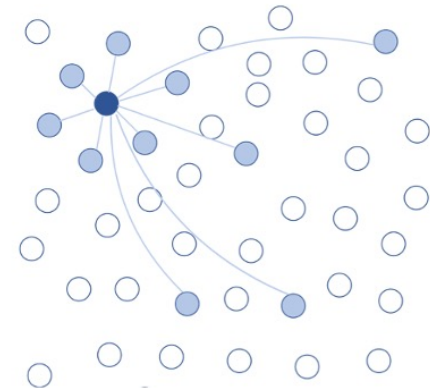
Information is broadcast directly from author to subscriber

Network



Information spreads from user to user connected through a network

Algorithm



Information is distributed by algorithm based on predictions of user preferences

Narayanan, 2023

Types of Information Propagation

	Subscription	Network	Algorithm
What User Sees			
Examples			
What determines reach			

Platforms rarely fall into one category completely, but helpful to think about

Narayanan, 2023

Types of Information Propagation

	Subscription	Network	Algorithm
What User Sees	Posts authored by those subscribe to	Posts authored and shared by those they've subscribed to	Posts algorithm predicts they will like best
Examples	Substack, newspapers, digital subscriptions, podcasts, original IG, original FB	Word of mouth, web, Twitter pre-2016, Mastodon	TikTok, YouTube
What determines reach	Poster's subscriber count	Subscriber count & content of the post	Content of the post

What model does your favorite social media use?

Subscriptions

- Subscription models – e.g. podcasts, cable and mainstream news – are understudied by **large proportion** of the information ecosystem
- **Television** is a major source of news – 5X time spent on TV compared to online (Allen et al, 2020)
- Americans spend **7 hours per week** listening to podcasts; an investigation found 1/10 of popular podcasts featured misinformation (Edison; Wirschafter, 2021)
- **No algorithm or network**
- Problem: Data Access

Case Studies:



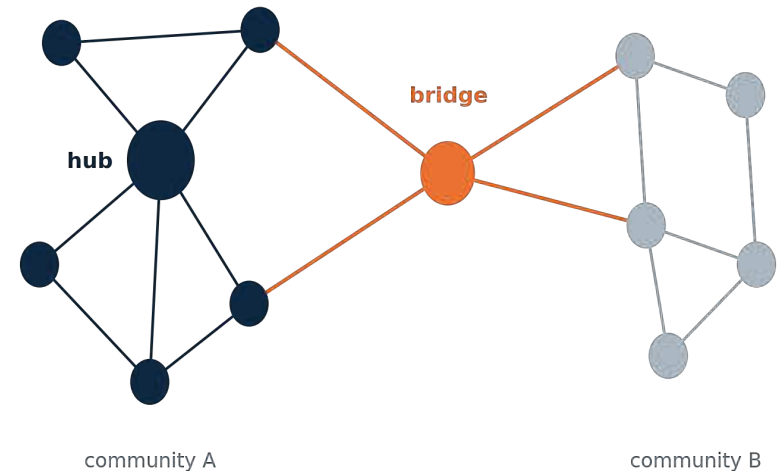
Joe Rogan



Local News

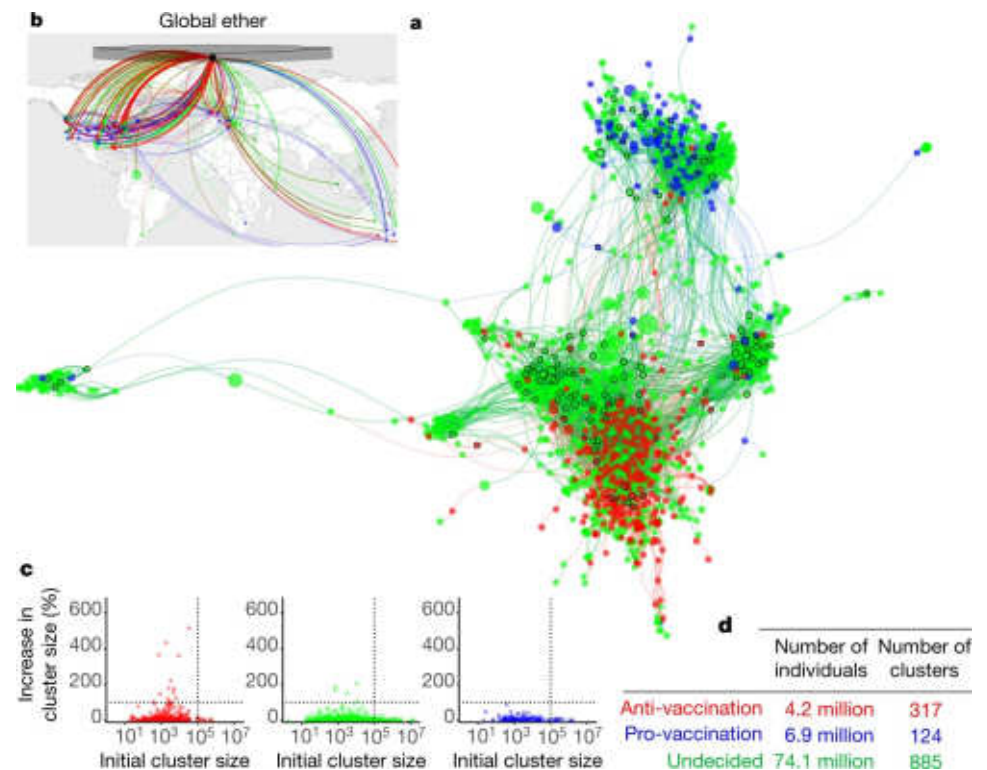
Networks

- Information spreads from account-to-account on social media through **cascades**, usually measured as chains of **reposts**, **reshares**, and **retweets**
- **Communities** = a densely connected cluster of accounts that mostly interact with each other
- **Hubs** = influential users with lots of connections (high **degree**)
- **Bridges** = an account or accounts linking otherwise-separate communities (high **betweenness**)



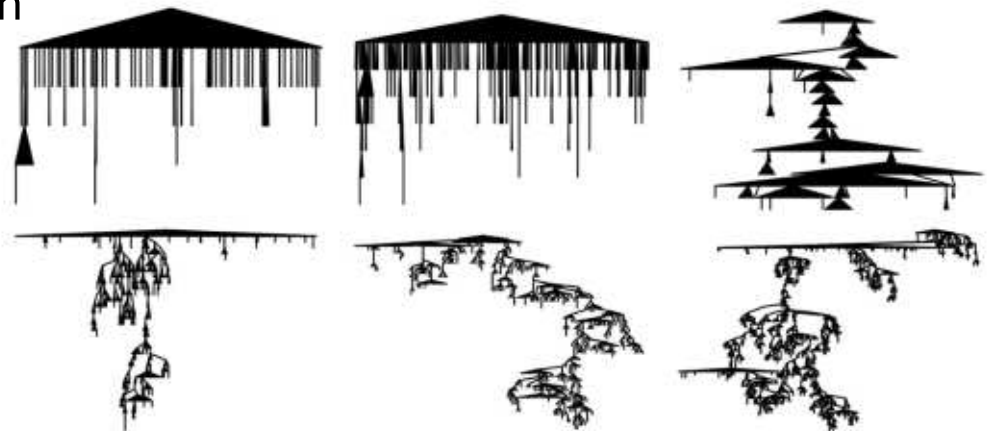
Networks – Vaccines

- Anti-vaccine and pro vaccine community are highly segregated (Johnson, 2020; Schmidt et al, 2018; Rathje et al, 2022)
- Anti-vax smaller, BUT had more connections / bridges to the “undecided” community and grew faster
- Pro-vaccine cluster was more of an **echo chamber**
- Risk: preaching to the choir



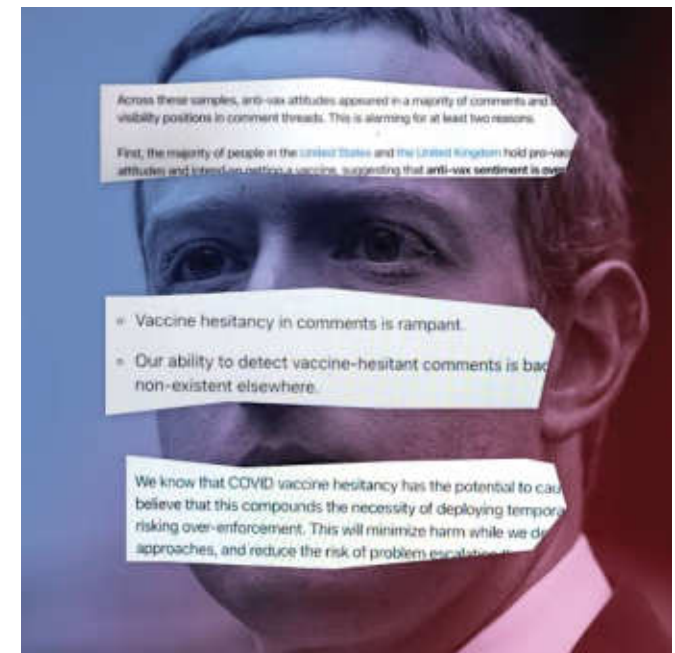
Networks - Virality

- Virality $\neq$ popularity
- “Going viral” = whether content spreads person to person in the network, rather than “broadcast”
- Very **unpredictable**
- Highly unequal $\rightarrow$ fewer than 1 in 100,000 tweets reach 1,000 retweets (Guinaudeau et al., 2022; Goel et al., 2016)
- Some evidence **that false information** is more likely to spread “virally” or via “deep reshares” (Facebook Files, Vosoughi et al, 2018)



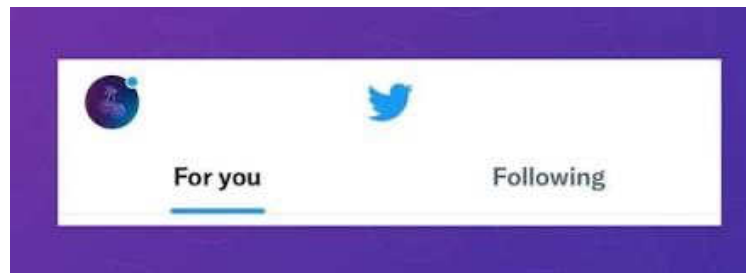
Case Study: Facebook Files, Covid-19 Vaccine

- Problematic vaccine content was **highly concentrated** — 111 accounts (0.016% of authors) drove 50% of anti-vaccine content views in the most affected communities (Facebook Files, 2021)
- The top 100 groups and 500 pages accounted for ~80% of downstream views on anti-vax content
- But much as 25-50% of vaccine content views and 50% of comments were vaccine-hesitant
- Spread from “core” (hubs) <-> through “inviters” (bridges) <-> to periphery
- Key risk: Inviters / Bridges



Algorithms

- Social network -> social *media*
- Recently, shift from **network model** to **algorithmic model**
- Platforms have more **discretion** over what content is distributed
- Goal: recommend content that is engaging

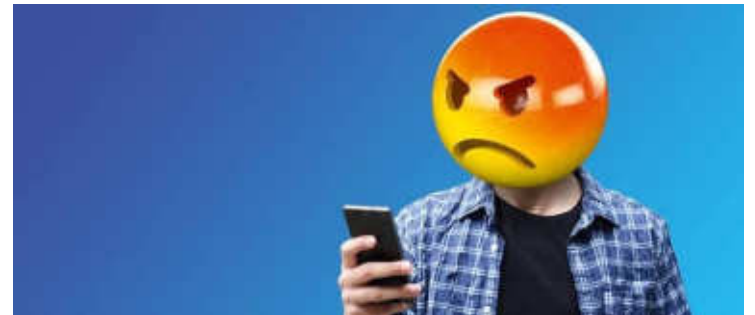


Algorithms: Engagement Prediction

- Engagement = the moment-to-moment **actions** of the user with **content** (watch, like, reshare, click)
- Goal: use features of the content to predict the user's likelihood of engagement with a post, given the past history of engagement with similar posts
 - You liked sports content in the past => show more sports content
- Platforms use **machine learning algorithms** to predict engagement
- Different platforms optimize for different types of engagement (e.g. likes/shares/watches on Facebook, view time on TikTok)
- Engagement → more time on platform → more money

What predicts engagement?

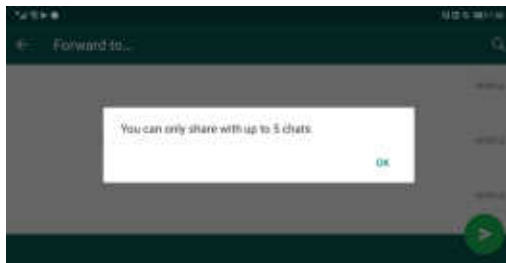
- It's complicated (and personalized)
- Evidence that content that is
 - Interesting (Berger & Milkman, 2012)
 - Negative (Robertson et al, 2020)
 - Novel (Vosoughi et al, 2018)
 - Toxic (Beknazar-Yuzbashev et al, 2026)
 - Morally outrageous (Brady et al, 2021)
- Attracts engagement on social media
- However, **high quality** sources get more **total views** than low quality sources (Allen et al, 2020; Guess et al, 2018; Guess et al 2021)



Vaccine Content & Engagement

- Some evidence that anti-vaccine **attracts more engagement** on social media than pro-vaccine content, although mixed by platform / time / vaccine (review - Wawrzuta et al, 2021)
- People still to **trustworthy news** outlets during pandemic, with low quality outlets capturing just 2.3% of web traffic and 14% of engagement (Altay et al, 2022)
- During pandemic, changing from engagement-ranking to trust-ranking decreased misinformation and increased for trustworthy sites (AP News, 2021)

Platform Actions

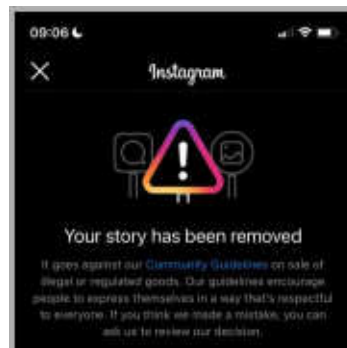


Friction “circuit breaker”



Algorithmic Demotion

Account Banning



Content removal



vacine autism

For You

Shop

Explore

Top

Users

Videos

LIVE

Photo



Autism information

Everyone's mental health story is unique. Empower yourself by learning about the autism spectrum from trusted sources. Learn more

Content promotion

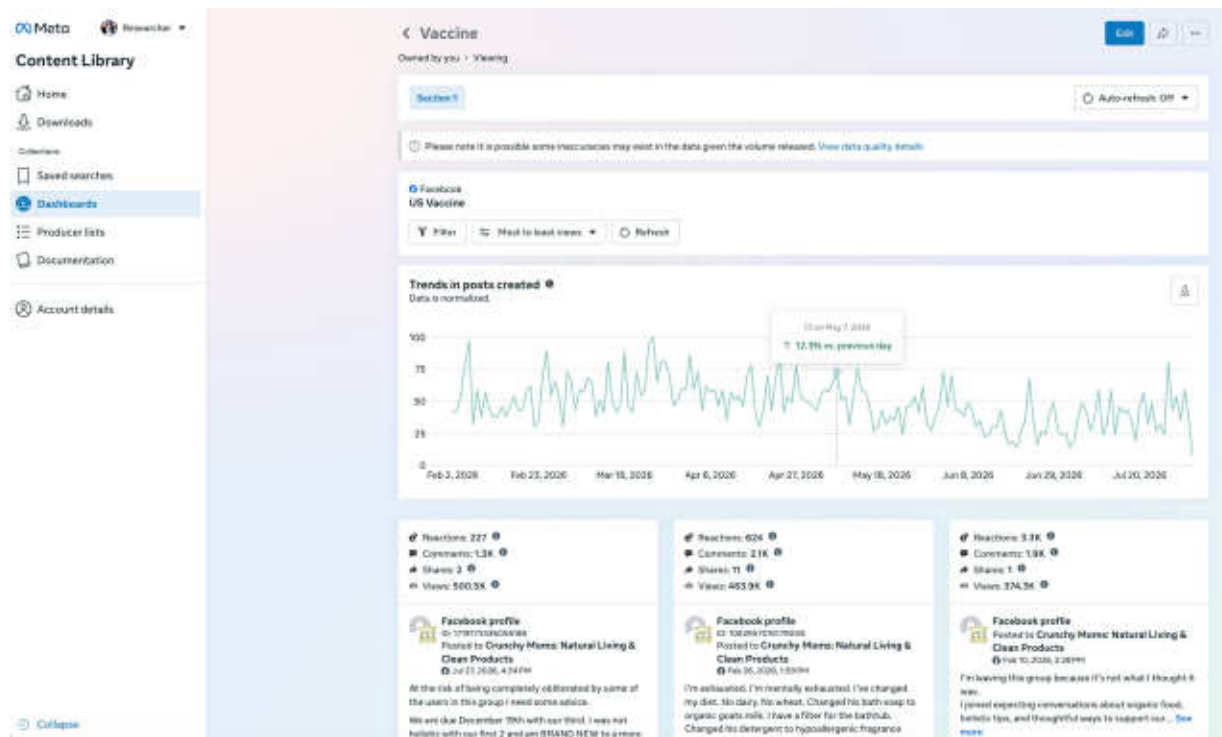
Think-Pair-Share

- Think (2 minutes)
 - Which platform actions (viral circuit breakers, demotion, banning) align with different content models? Are there any other actions?
 - What are the downsides or risk factors of each of these methods?
- Pair (4 minutes)
 - Share with the person next to you, and discuss what the major risk factors are
- Share (4 minutes)
 - Discuss

Challenges: Data Access

- **No longer accessible**
 - **CrowdTangle** (Facebook/Instagram public posts) — shut down Aug 2024
 - **Twitter/X academic API** — free tier ended 2023; enterprise access now runs thousands per month
 - **Reddit Pushshift** — restricted to moderators, 2023
- **Status quo**
 - **Meta Content Library** — public FB/IG/Threads archive, but: approval required, no raw download, analysis inside a virtual clean room, retrieval caps (~500k records/week)
 - **TikTok Research API** — US/EU academics only; documented coverage gaps and reproducibility problems
 - **YouTube Data API** — quota-limited
- **Potential future: DSA article 40**

Meta Content Library



Break

Activity: Vaccine Policy Red Team

1. Go through assignment & split up into 2 teams (5 mins)
2. Write a vaccine information policy for a multinational social media company (15 mins)
3. Share policies with group (10 mins)
4. Brainstorm cases that would “break” the other team policy (15 mins)
5. Share “red team” examples (15 mins)
6. Final discussion + Wrap Up (15 mins)

Bookface

Monthly active users	2.1 billion
Countries of operation	180
Interface languages	41
Content posted daily	~4 billion items
Revenue	Advertising, 97% of total
Revenue driver	Impressions, which scale with time on platform and total users

Content Distribution model: Algorithm and Network

1	Content similar to what the user has liked in the past
2	Content shared and liked by the user's connections and groups
3	Content that is popular elsewhere on the site

Bookface's Goals

1	Maintain and increase engagement	Revenue is proportional to the amount of content people see and the amount of time they spend on the platform
2	Legal Compliance	- EU Digital Services Act obligations says that company must remove false information - US state have introduced statutes viewpoint-based moderation. - There is active litigation asserting both too much and too little.
3	Promote accurate health information and limit health misinformation	The company wants to promote accurate public health information and remove misinformation, while preserving the platform as a public forum for debate

Resources

Contracted reviewers	15,000 across 20 sites, mostly outside the markets they review
Median time per item	15 seconds
Daily review capacity	~1 million items
Content posted daily	~4 billion items
First-pass triage	Automated classifiers, materially better in English than elsewhere
Third-party fact-checking	60 languages; a few hundred ratings per market per week; 1–3 day lag

Round 1: Write the policy

1. Write a list of **3-5** rules targeting vaccine and health content
2. Each rule should specify the category of content and the action: promote, no action, label, demote, or remove.
3. For each rule, how is content identified? Automated classifier, experts, user report plus review, source, account-level designation, or a hybrid. Consider the resources you have.

Discuss round 1

Round 2: Red Team

1. Come up with a list of 5 examples that **test the policy**
2. Example positions
 1. You are an anti vaccine advocacy group. What strategies could you use to evade the policy? Is there any content that might evade detection and be promoted by the policy? Consider what we learned about what content attracts engagement on social media, and the network structure of pro and anti vaccine communities
 2. You are a public health organization. Is there any content that might be actioned by the policy that you would promote?
 3. You are a politician. How might you attack the policy as being contrary to your political views?
 4. Is there a recent public vaccine crisis that would **test** the policy?

Discuss round 2

Other examples

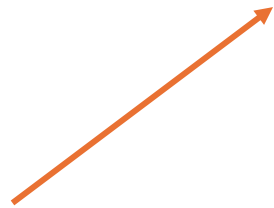
- A parent's account of her son's febrile seizure after his MMR shot
- A news report: Chinese vaccine maker fined \$1.3bn for falsified records
- “Has anyone else's baby run a fever for two days after the 6-month shots?”
- A cardiologist explaining myocarditis rates in young men after mRNA vaccines
- A parent saying she's spacing out her second child's schedule
- “The vaccine is 100% effective. You cannot catch COVID or give it to anyone.”
- CDC guidance shifts and recommends different vaccine schedule
- Anti vaccine advocates coordinated comment disagreement on promoted pro-vaccine posts

Debrief

Case Study: Covid-19 Vaccine Information on Facebook

Impact

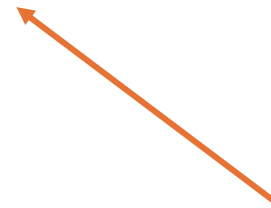
Does social media content impact vaccine hesitancy?
Which content has the biggest impact on vaccine hesitancy?



Reach

What kind of vaccine content do people see
on social media?

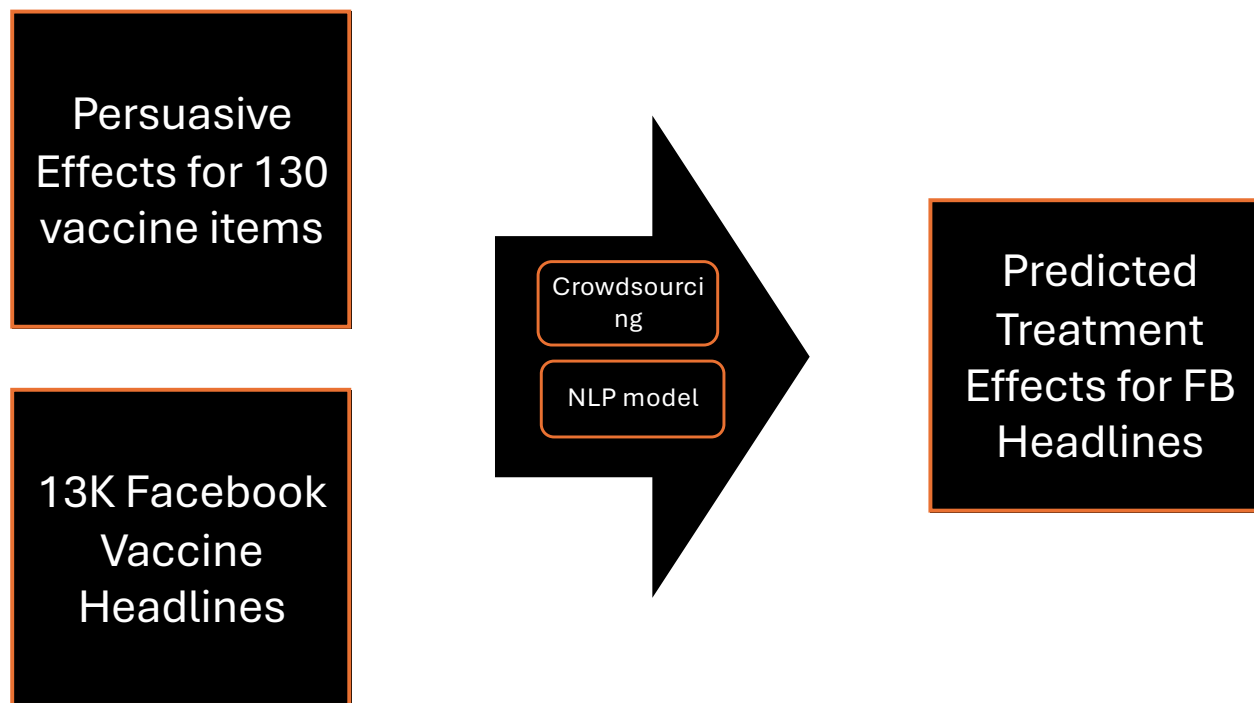
Who does it reach?



Persuasive Effect

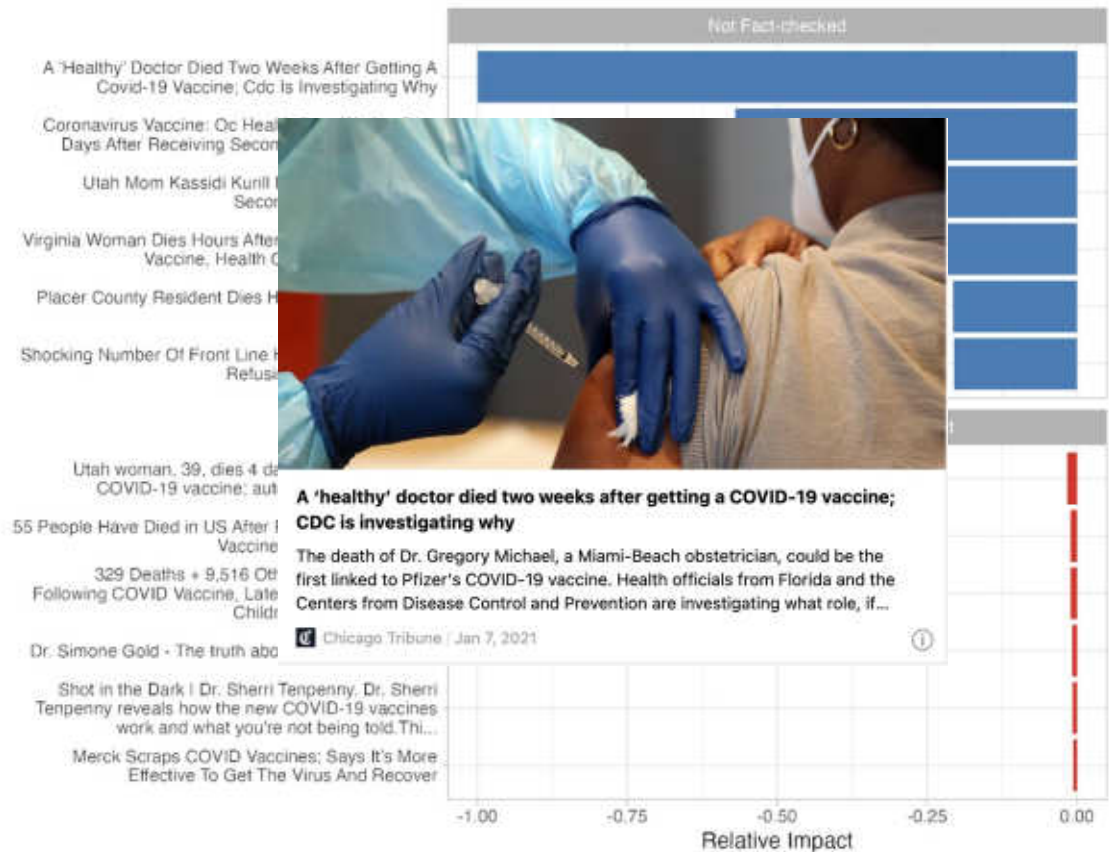
Which types of vaccine information influences
vaccine hesitancy?

Combine Persuasive Effects & Exposure Data



Impact Estimates

The impact of gray area vaccine-skeptical content is **50X** that of clear misinformation

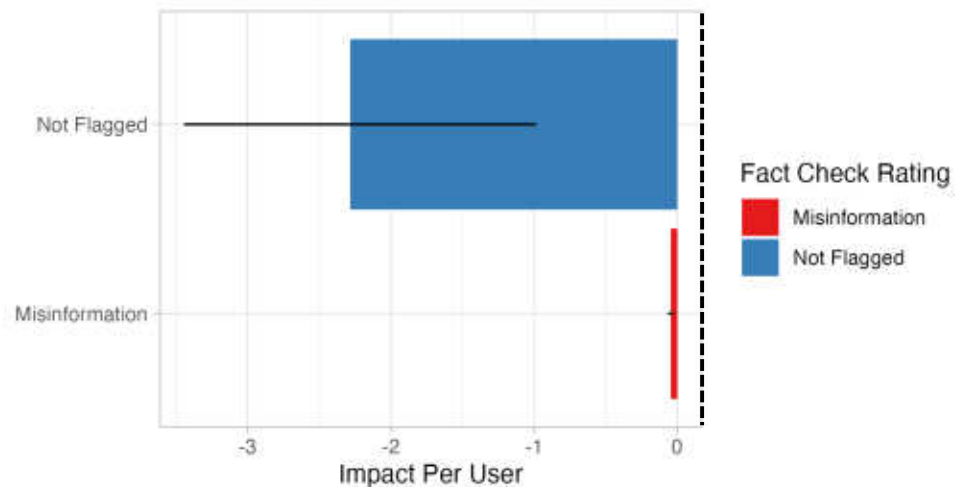


Aggregate Impact of Misinformation vs. Vaccine-Skeptical Content

Misinformation content lowers vaccine intentions by predicted **.05pp (0.07, .02)** per user

Vaccine-skeptical content lowered vaccine intentions by predicted **2.28pp (3.4, 1)** per user

A **~50X** difference!



Top Stories, Conditional on Exposure



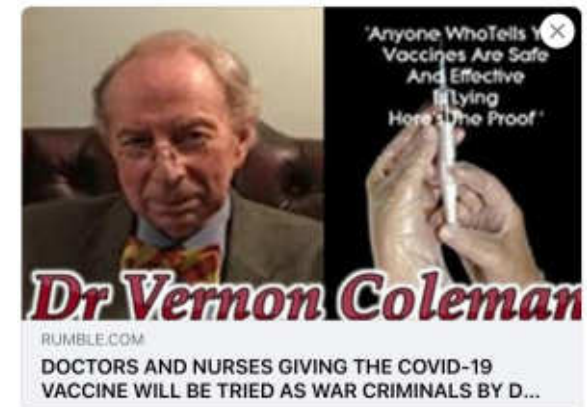
ALP, CHINA, CHINESE CONTROL, COVID COPS, COVID VACCINES, COVID-19, GENERAL, LIBERAL NATIONAL PARTY, VACCINES

**Not one person with a brain
would take a Covid test or
vaccine after watching this
disturbing video produced by
doctors**



**Shocking! Miscarriages Increase by 366% in Just 6 Weeks Because of
Covid Vaccines**

Humanity's Be Free



Top Stories, Weighted by Exposure



Health care worker dies after second dose of COVID vaccine, investigations underway

Tim Zook, a resident of Orange who worked at an acute care medical center, fell ill within hours of his second Pfizer-BioNTech shot. He spiraled downward and died four days later.

Orange County Register · Teri Sforza · Jul 15, 2021



A 'healthy' doctor died two weeks after getting a COVID-19 vaccine; CDC is investigating why

The death of Dr. Gregory Michael, a Miami-Beach obstetrician, could be the first linked to Pfizer's COVID-19 vaccine. Health officials from Florida and the Centers for Disease Control and Prevention are investigating what role, if...

Chicago Tribune · Jan 7, 2021



Utah woman, 39, dies 4 days after 2nd dose of COVID-19 vaccine; autopsy now completed

SALT LAKE CITY (KUTV) — During a KUTV investigation into COVID-19 vaccine side effects and where to report them, we found four reported deaths, filed by Utah families and their caregivers to the CDC's Vaccine Adverse Effects Reporting...

KMPH · HEDD HATCH · MAR 10, 2021



>50 million people viewed
9 million = all misinformation

File Edit View Insert Format Tools Add-ons Help Last edit was made 2 days ago by 100% Heading 1 Arial 20 B I U A 1 2 3 4 5 6 7				
Approved Product Enforcement Mapped to Tiers:				
Tier	Description	Content-Level	Dedicated Entities	Comments
Tier 0 - IS Violating (Prevalence Reduction Goal)	Health Misinfo (RFH, WDH, M&H)	FB+IG: Remove at scale		
	Explicit CV19 Vaccine Discouragement & Interference	FB+IG: Remove (on_esc)		
Tier 1 - Alarmism & Criticism (Prevalence Reduction Goal)	Sensational or Alarmist Vaccine Content	FB: Demote, Reduce in Search, no notifs, non_rec	FB: Demote, Reduce in Search, no notifs, non_rec IG: Tier 2 non_rec, exact match in search	FB+IG: Demote comments AND Filter from preview/inline
	Criticizing Choice to Receive/Provide Vaccines	IG: non_rec		
Tier 2 - Indirect Discouragement (Push/search/virality Goal)	Promoting Vaccine Refusals & Alternatives	FB+IG: Non-Rec + Inform/Friction	FB: Demote, Reduce in Search, no notifs, no rec, AND inform IG: Tier 2 non_rec, exact match in search	
	Shocking Stories			
Tier 3 - Objections & Skepticism (Legitimacy + KAP Goal)	Objections & Skepticism: Personal choice & beliefs	FB+IG: Inform/Friction (selected viral and salient subtopics)	FB+IG: Non-Rec AND Inform/Friction	FB+IG: Filter from preview/inline
	Objections & Skepticism: Institutions & development			
Tier 4 - Neutral	Other Vaccine Discussion & Debate		No action	No action
5/3/21 Note: [UPDATE] As a break glass measure, we will be demoting both Tier 1 and Tier 2 content at 50% strength when manually reviewed				

Not just a Facebook problem

