

Fighting Disinformation: Another Noble Experiment?

David Beaver (The University of Texas at Austin)

Jason Stanley (Yale University)

1

Plan

- A noble experiment
- Good Systems
- Supply and demand
- Psych 101
- Fact check this!
- Closing thoughts

2

A noble experiment

3

ALL MERCHANDISE ADVERTISED IN THE TRIBUNE IS GUARANTEED

New York Tribune

First to Last - the Truth: News - Editorials - Advertisements

WEDNESDAY, DECEMBER 18, 1917

WEATHER: Partly cloudy today and tomorrow with rising temperature; moderate southerly winds.

Vol. LXXVII No. 25,963

Copyright, 1917, by The Tribune Publishing Co.

Price, Five Cents

Published by The Tribune Publishing Co., 410 N. W. Cor. of 42d St. and 7th Ave., New York, N. Y.

Kuehlmann on Way to Join Peace Party

German Foreign Minister Reported Proceeding to Brest-Litovsk

Bolshviki Beaten in Odessa Fighting

Sulfur Divides Forces, and Many Are Killed; Kaldine Not Captured

Hertling Declares Just Peace With Britain Impossible

German Chancellor Asserts Speech of Lloyd George Has Made Idea Useless

French Votes Go to Laurier

Soldiers' Riots Expected to Swell Vote for Premier

Draft Wins In Canada By Big Majority

Borden and Union Government Returned for Three More Years

French Votes Go to Laurier

Soldiers' Riots Expected to Swell Vote for Premier

Germans Raid British Convoy, Sink 11 Ships

British Destroyer and Four Mine Sweepers Lost With Six Merchants

Second Warship Hit In North Sea Attack

Vessels Prepared Only for Submarines; Investigation Begun

House Adopts Prohibition Amendment by 282 to 128

Measure Goes to Conference - Period for Ratification Is Only Difference

Parties Evenly Divided on Vote

141 Democrats and 137 Republicans Join Opponents to Vote "Yes"

Morgan Gives Art Treasures To Museum

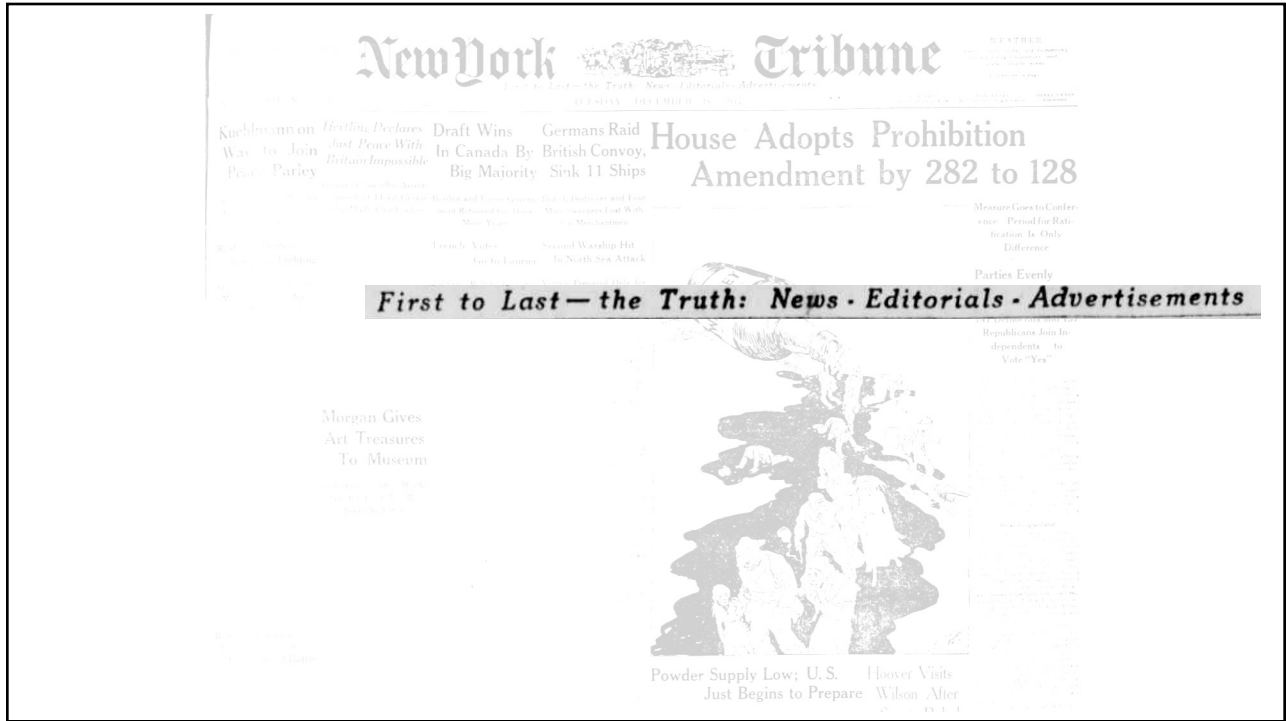
Collection of 3,000 Works Greater Than Any Bestowed by Father

Powder Supply Low; U. S. Just Begins to Prepare

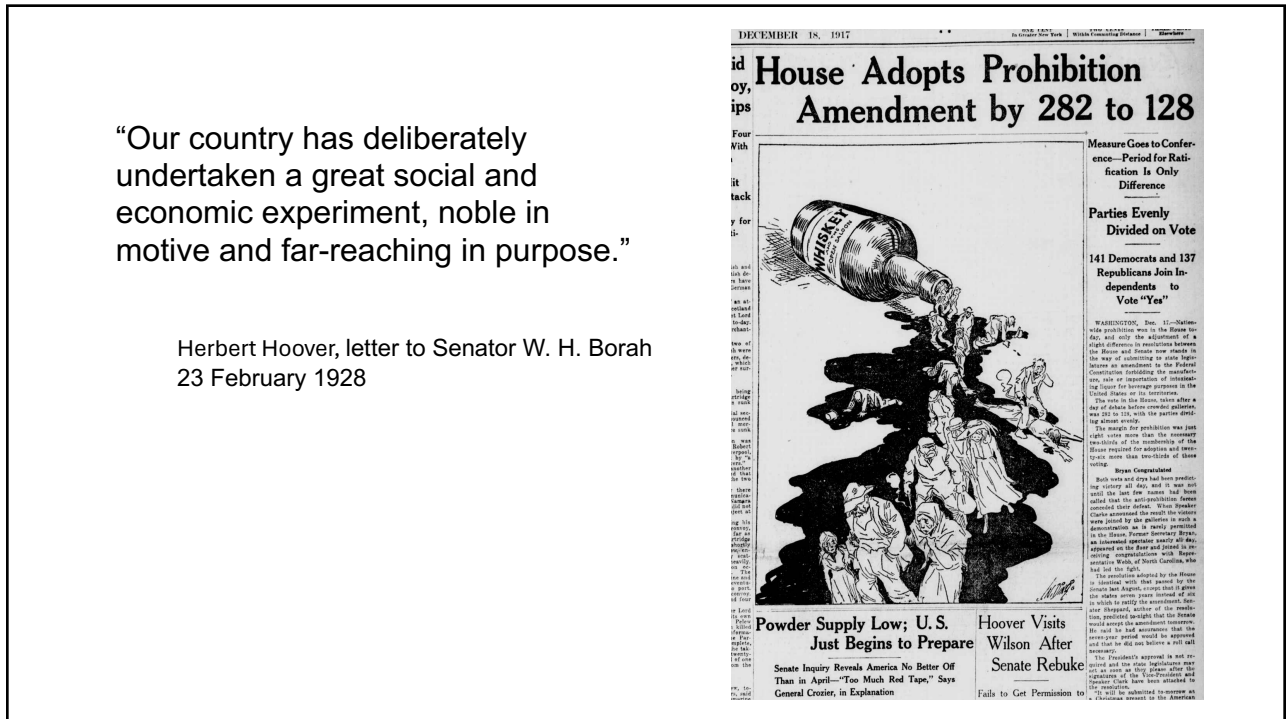
Hoover Visits Wilson After

WASHINGTON, Dec. 17.—The House today voted 282 to 128 to pass the prohibition amendment to the constitution. The measure will go to conference with the Senate. The amendment provides for the prohibition of the manufacture and sale of intoxicating liquors within the United States or its territories. The vote was evenly divided on the question of ratification, with 141 Democrats and 137 Republicans joining the opponents to vote "yes".

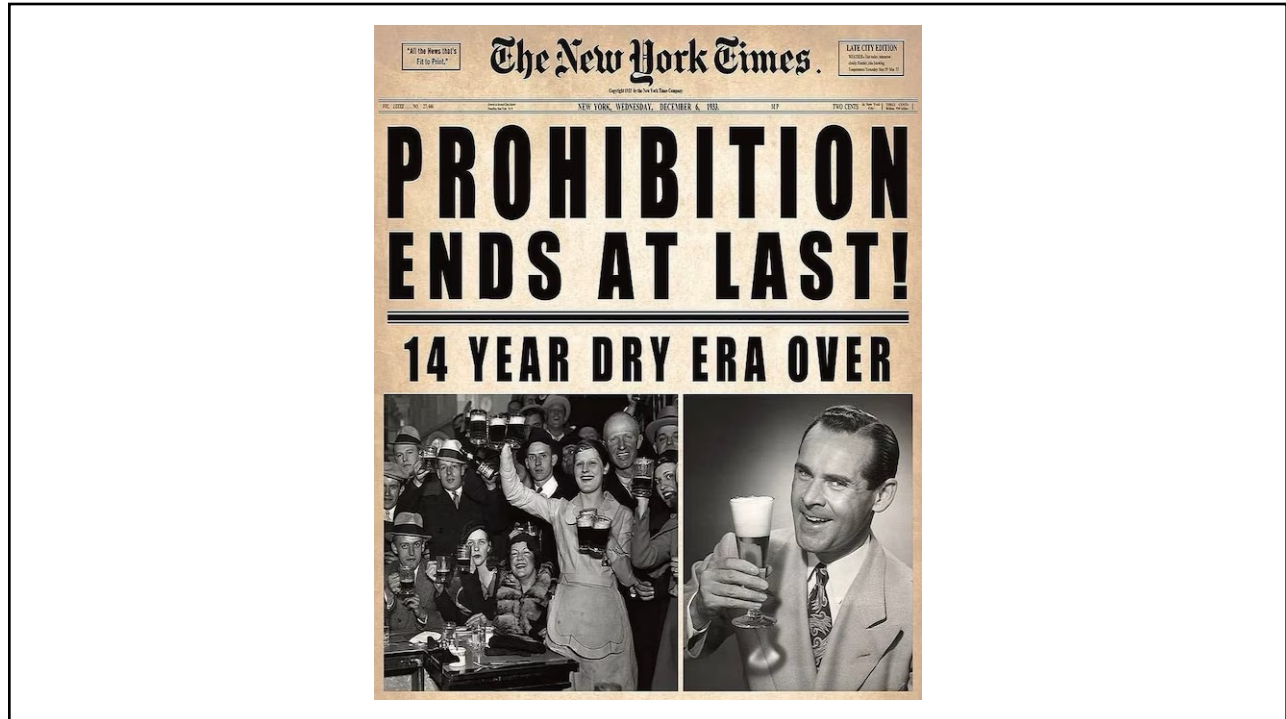
4



5



6



7



8

A new social experiment?

What if we could fact check the entire web, all social and traditional media, and perhaps everything everyone ever said to you on a phone call or in a zoom meeting?

Wouldn't it be great to be able to filter out every false claim?

9

Good Systems

10

COMMUNICATIONS
of the
ACM


Explore Topics ▾ Latest Issue ▾ Q Sign In

OPINION Artificial Intelligence and Machine Learning

Now, Later, and Lasting: 10 Priorities for AI Research, Policy, and Practice

Shaping the future of artificial intelligence.

By Eric Horvitz, Vincent Conitzer, Sheila McIlraith, and Peter Stone
Posted May 1 2024



“Disinformation. Address uses of AI by malicious actors for disinformation, manipulation, and impersonation, which pose a risk to an informed citizenry and to the fabric of our democracies. Beyond harnessing promising technical and sociotechnical strategies built on content provenance, watermarking, and fingerprinting, investigate new laws, regulations, and enforcement to combat these threats while protecting civil liberties. The goal is to safeguard the integrity of information and communication against malevolent repurposing of widely available AI methods to deceive and persuade.”

11

COMMUNICATIONS
of the
ACM


Explore Topics ▾ Latest Issue ▾ Q Sign In

OPINION Artificial Intelligence and Machine Learning

Now, Later, and Lasting: 10 Priorities for AI Research, Policy, and Practice

Shaping the future of artificial intelligence.

By Eric Horvitz, Vincent Conitzer, Sheila McIlraith, and Peter Stone
Posted May 1 2024



“Disinformation. Address uses of AI by malicious actors for disinformation, manipulation, and impersonation, which pose a risk to an informed citizenry and to the fabric of our democracies. Beyond harnessing promising technical and sociotechnical strategies built on content provenance, watermarking, and fingerprinting, investigate new laws, regulations, and enforcement to combat these threats while protecting civil liberties. The goal is to safeguard the integrity of information and communication against malevolent repurposing of widely available AI methods to deceive and persuade.”

12

GOOD SYSTEMS

Ethical AI at UT Austin

“

How can we ensure that AI is beneficial — not detrimental — to humanity? What unintended consequences are we overlooking by developing technology that can be manipulated and misused?

It is ethically irresponsible to focus only on what AI can do. We believe it is equally important to ask what it should (and should not) do.

Our goal is to better understand what changes new technologies will bring, predict how those changes will unfold, and mitigate the harms or unintended consequences they could cause while still leveraging the benefits AI provides.

”

13

Moderating Social Media Discourse for a Healthy Democracy

Dr. Josephine Lukito, Kathryn Kazanas, Bin Chen

Policy Brief, Published October 26, 2022
Designing Responsible AI Technologies to Curb Disinformation Project

To address this problem space, we advocate for the following solutions:

1. **Algorithmic governance & transparency:** Internet Service Providers should be more transparent with users about content moderation policies and algorithms, and clarify users' basic rights on the platform.
2. **Flagging recommendations:** We advocate a more effective, efficient and comprehensive flagging system through a combined strategy of content- and user-based approaches.
3. **Multiplatform collaboration:** Fighting harmful online content requires a collaborative effort among policy makers, civil society groups, researchers, and different platforms.
4. **Long-term considerations:** Building a regular and prolonged tracking system is essential to make anti-misinformation efforts more efficient and effective, especially in complex scenarios.

14

Supply and demand

15

A toxicity metaphor

- Disinformation is a pollutant or disease.
 - Our information space is being contaminated.
 - The toxicity of disinformation sickens the public and harms democracy.
-
- Inoculate the public & provide them with “masks”!
 - Clean the air!
 - Penalize or eliminate the polluters!

16

An addiction metaphor

- People are addicted to disinformation.
 - Disinformation creators and traffickers are giving people what they want.
 - The epidemic of disinformation addiction stems from underlying conditions: socioeconomic distress, voicelessness, alienation.
-
- ~~Provide the public with masks~~
 - ~~Clean the air~~
 - ~~Eliminate the polluters~~

17

Why is there a thirst for disinformation?

18

Why is there a thirst for disinformation?

- There isn't.

19

Psych 101

20

~~Psych 101~~ PSY 301

21

The problem with ideal humans

- An enlightenment approach suggests that man is a rational animal, a curious and open-minded seeker of truth.
- A century of psychology suggests otherwise.
- We have **blindspots** to our own irrationality, and form beliefs according to emotional and social need.
- When experiencing extreme **cognitive dissonance**, the emotional tail wags the reasoning dog.
- People facing dissonance seek social connection, information and ideas that **cohere**.
- People resist painful inferences and show **confirmation bias**.
- *When suffering, you seek, listen to and accept only what helps you.*

22

The social side

- Individual psychology does not suffice to understand the dynamics of disinformation addiction.
- Festinger's classic studies of **cults**, and more recent social psychology build on an important tradition of socio-anthropological work (Gustave Le Bon, Gabriel Tarde, Durkheim)
- We deal with dissonance by strengthening ties to an in-group, and "explaining" problems to be created by out-groups.
 - Groups turn inward, developing their own practices, their own languages, and their own "facts."

23

Why is there a thirst for disinformation?

- There isn't.
- There is a thirst for social connection and convenient truths.

24

Why is there a thirst for disinformation?

- There isn't.
- There is a thirst for social connection and convenient truths.
- Both social media companies and politicians greedily exploit these *Us-them* tendencies.

25

Fact check this!

26

Symbolism Will Be Their Downfall

Their need for symbolism will be their downfall.
Follow the Owl & Y head around the world.
Identify and list.
They don't hide it.
They don't fear you.
You are sheep to them.
You are feeders.
Godfather III.

- Qdrop 11/21/2017

27



Donald J. Trump ✓
@realDonaldTrump

Follow

There is a Revolution going on in California. Soooo many Sanctuary areas want OUT of this ridiculous, crime infested & breeding concept. Jerry Brown is trying to back out of the National Guard at the Border, but the people of the State are not happy. Want Security & Safety NOW!

2:59 AM - 18 Apr 2018

28,857 Retweets 113,559 Likes



28

Rep. Marjorie Taylor Greene @RepMTG

Antisemitism is wrong, but I will not be voting for the Antisemitism Awareness Act of 2023 (H.R. 6090) today that could convict Christians of antisemitism for believing the Gospel that says Jesus was handed over to Herod to be crucified by the Jews.

Read the bill text and contemporary examples of antisemitism like #9.

whole-of-government-and-society approach.

4. DEFINITIONS.

For purposes of this Act, the term "definition of anti-Semitism"—

(1) means the definition of antisemitism adopted on May 26, 2016, by the IHRA, of which the United States is a member, which definition has been adopted by the Department of State; and

(2) includes the "[c]ontemporary examples of antisemitism" identified in the IHRA definition.

"Antisemitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of antisemitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities."

Contemporary examples of antisemitism in public life, the media, schools, the workplace, and in the religious sphere could, taking into account the overall context, include, but are not limited to:

1. Calling for, aiding, or justifying the killing or harming of Jews in the name of a radical ideology or an extremist view of religion.
2. Making mendacious, dehumanizing, demonizing, or stereotypical allegations about Jews as such or the power of Jews as collectives — not as individuals but not exclusively, the myth about a world Jewish conspiracy or of Jews controlling the media, economy, government or other societal institutions.
3. Accusing Jews as a people of being responsible for real or imagined wrongdoing committed by a single Jewish person or group, or even for acts committed by non-Jews.
4. Denying the fact, scope, mechanisms (e.g. gas chambers) or intentionality of the genocide of the Jewish people at the hands of National Socialist Germany and its supporters and accomplices during World War II (the Holocaust).
5. Accusing the Jews as a people, or Israel as a state, of terrorism or other forms of violence, or of practicing racial discrimination or incitement to violence, when the actions of the Jewish people or the state of Israel are in fact the victims of such violence or incitement.
6. Denying Jewish claims of being more persecuted in the world than other persecuted groups of Jews worldwide, done in the interests of their own nation.
7. Denying the Jewish people their right to self-determination, e.g., by claiming that the existence of a State of Israel is a racist endeavor.
8. Applying double standards by requiring of a Jew behavior not expected or demanded of any other democratic citizen.
9. Using the symbols and images associated with classic antisemitism (e.g., claims of Jews killing Jesus or blood libel) to characterize Israel or Israelis.
10. Drawing comparisons of contemporary Israeli policy to that of the Nazis.
11. Holding Jews collectively responsible for actions of the state of Israel.

Last edited 2:38 PM · May 1, 2024 · **820.8K** Views

2K 2.7K 7.2K 548

29

WASHINGTON, May 25 (Reuters) - U.S. Representative Marjorie Taylor Greene, a firebrand ally of former President Donald Trump, was broadly denounced by her fellow Republicans on Tuesday for likening COVID-19 masks and vaccinations to the Nazi Holocaust against the Jews.

30

“Your conduct is in violation of Penal Code Sections 42.01 Disorderly Conduct, 42.02 Riot, 42.03 Obstructing a Highway or other passageway, 30.05 Criminal Trespass. I command you in the name of the University of Texas Police Department to disperse, and if you do not, you shall be arrested for Violation of Penal code Section 42.01 Disorderly Conduct, 42.02 Riot, 42.03 Obstructing a Highway or other passageway, 30.05 Criminal Trespass.”

- UTPD, 4/29/2024, Message to UT community

31

“I’m thankful we live in a country where free expression is a fiercely protected Constitutional right. I’m grateful that our campus has seen 13 pro-Palestinian events take place during the past several months largely without incident — plus another one today. I am grateful that everyone is safe after yesterday, we continue to hold in-person classes, and that today’s events followed our long-standing campus standards for allowed demonstrations. And I’m grateful to work in a place where people with different points of view come together to try to solve hard problems.”

- Pres. Jay Hartzell, 4/25/2024

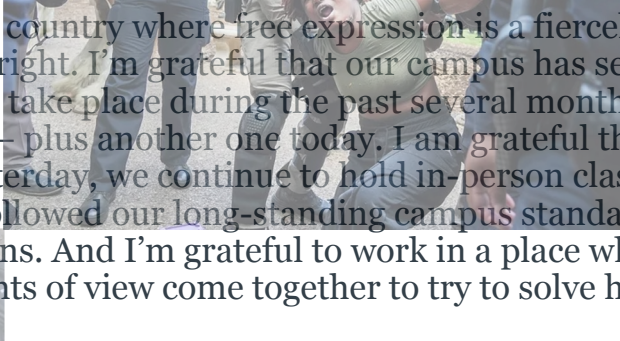
32

We will continue to safeguard the free speech and assembly rights of everyone on our campus, while we protect our University and students, who are preparing for their final exams.

Sincerely,
Internal Communications

- message to UT community, 4/30/2024

33



“I’m thankful we live in a country where free expression is a fiercely protected Constitutional right. I’m grateful that our campus has seen 13 pro-Palestinian events take place during the past several months largely without incident — plus another one today. I am grateful that everyone is safe after yesterday, we continue to hold in-person classes, and that today’s events followed our long-standing campus standards for allowed demonstrations. And I’m grateful to work in a place where people with different points of view come together to try to solve hard problems.”

- Pres. Jay Hartzell, 4/25/2024

Images: Austin American-Statesman 4/24/2024

34



35

Greg Abbott @GregAbbott_TX · Apr 24
Arrests being made right now & will continue until the crowd disperses.

These protesters belong in jail.

Antisemitism will not be tolerated in Texas. Period.

Students joining in hate-filled, antisemitic protests at any public college or university in Texas should be expelled.

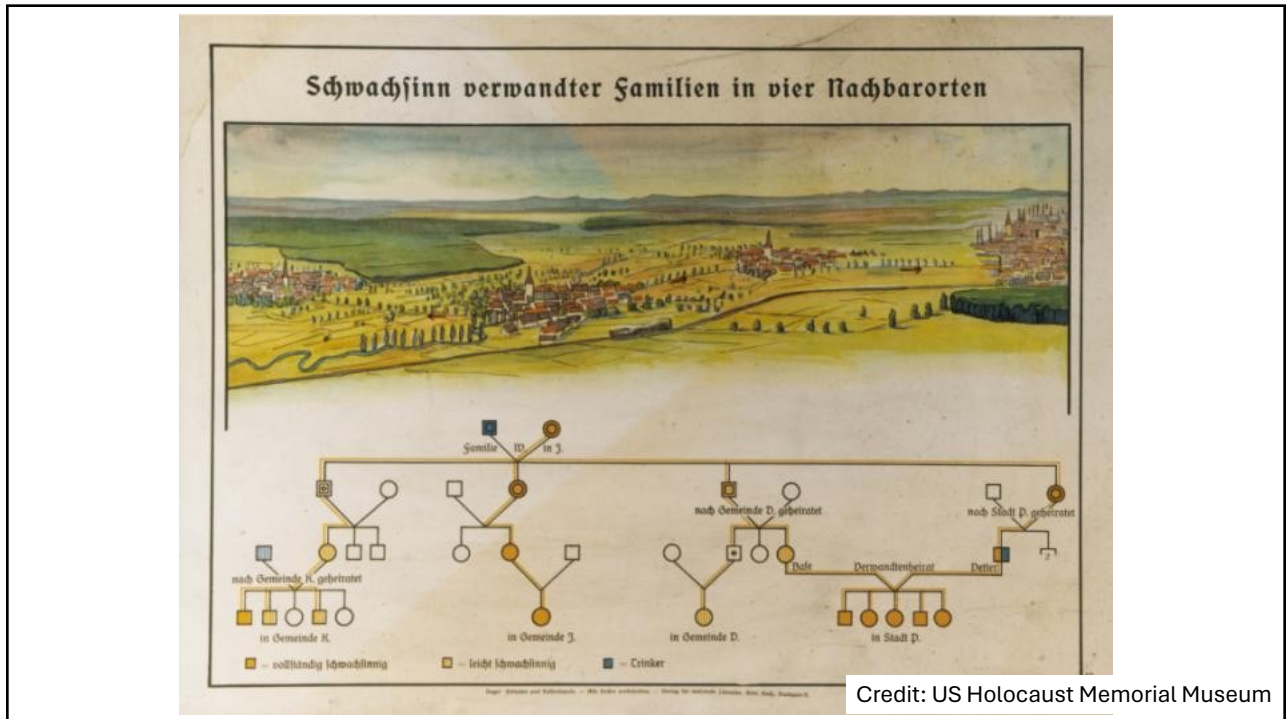
Greg Price @greg_price11 · Apr 24
Pro Hamas idiots at UT Austin discover what happens why you try to pull a Columbia in Texas

A photograph of a large crowd of people gathered in front of a brick building with arched windows. A banner in the foreground reads "TEXAS FIGHT". Many people are wearing military-style gear, including helmets and vests.

36



37



Credit: US Holocaust Memorial Museum

38

Closing thoughts

39

Concerns with tools for detecting disinformation:

- Propaganda is a jigsaw of emotional appeal, social differentiation and selective reporting. Disinformation is a small part of that puzzle and removing it would change little.
- People seek convenient facts and won't use tools that seem to censor what they approve of.
- People in power like to spread their own "truths" and will avoid tools they don't like.
- Worse, they will subvert those tools by basing them on their own version of ground truth.

40

Conclusion

- The goal of building “good systems” is worthy, but we need to be very careful in determining *who* they are good for.
- Anything we would collectively agree is an improvement in our information spaces must depend on a collective will to improve them.
- Psychology, social dynamics, economic and political incentives are at present aligned in favor of disinformation and not against it.
- Disinformation is not just a matter of supply, but also demand.
- Building “good systems” to clean up the information space while ignoring this is potentially just another ***noble experiment***.

41

Conclusion

- The goal of building “good systems” is worthy, but we need to be very careful in determining *who* they are good for.
- Anything we would collectively agree is an improvement in our information spaces must depend on a collective will to improve them.
- Psychology, social dynamics, economic and political incentives are at present aligned in favor of disinformation and not against it.
- Disinformation is not just a matter of supply, but also demand.
- Building “good systems” to clean up the information space while ignoring this is potentially just another ***noble experiment***.
- Final worry: **claims of neutrality are always assertions of power.**



42