

An Agenda for Disinformation Research

A Computing Community Consortium (CCC) Quadrennial Paper

Joshua Garland

Santa Fe Institute

Joshua@santafe.edu





Nadya Bliss
Arizona State University
Center on Narrative, Disinformation,
and Strategic Influence
CCC



Scott W. Ruston
Arizona State University
Center on Narrative, Disinformation,
and Strategic Influence



Elizabeth Bradley
University of Colorado, Boulder
Santa Fe Institute
CCC



Kate Starbird
University of Washington



Filippo Menczer
Indiana University



Chris Wiggins
Columbia University
New York Times

COMPUTING COMMUNITY CONSORTIUM



The CCC's mission is to enable the pursuit of innovative, high-impact research that aligns with pressing national and global challenges

- CCC is of, by, and for the computing research community: a responsive, respected, visionary organization that seeks diversity, equity, and inclusivity in all of its activities.
- CCC is a powerful convener that brings together thought leaders from industry, academia, and government to articulate and advance compelling research visions.
- CCC is an effective communicator with stakeholders, policymakers, the public, and the broad computing research community regarding the substance, and the importance, of those visions.



CCC

Computing Community Consortium
Catalyst

What is disinformation?

False, misleading or inaccurate information disseminated with the *intent to deceive*

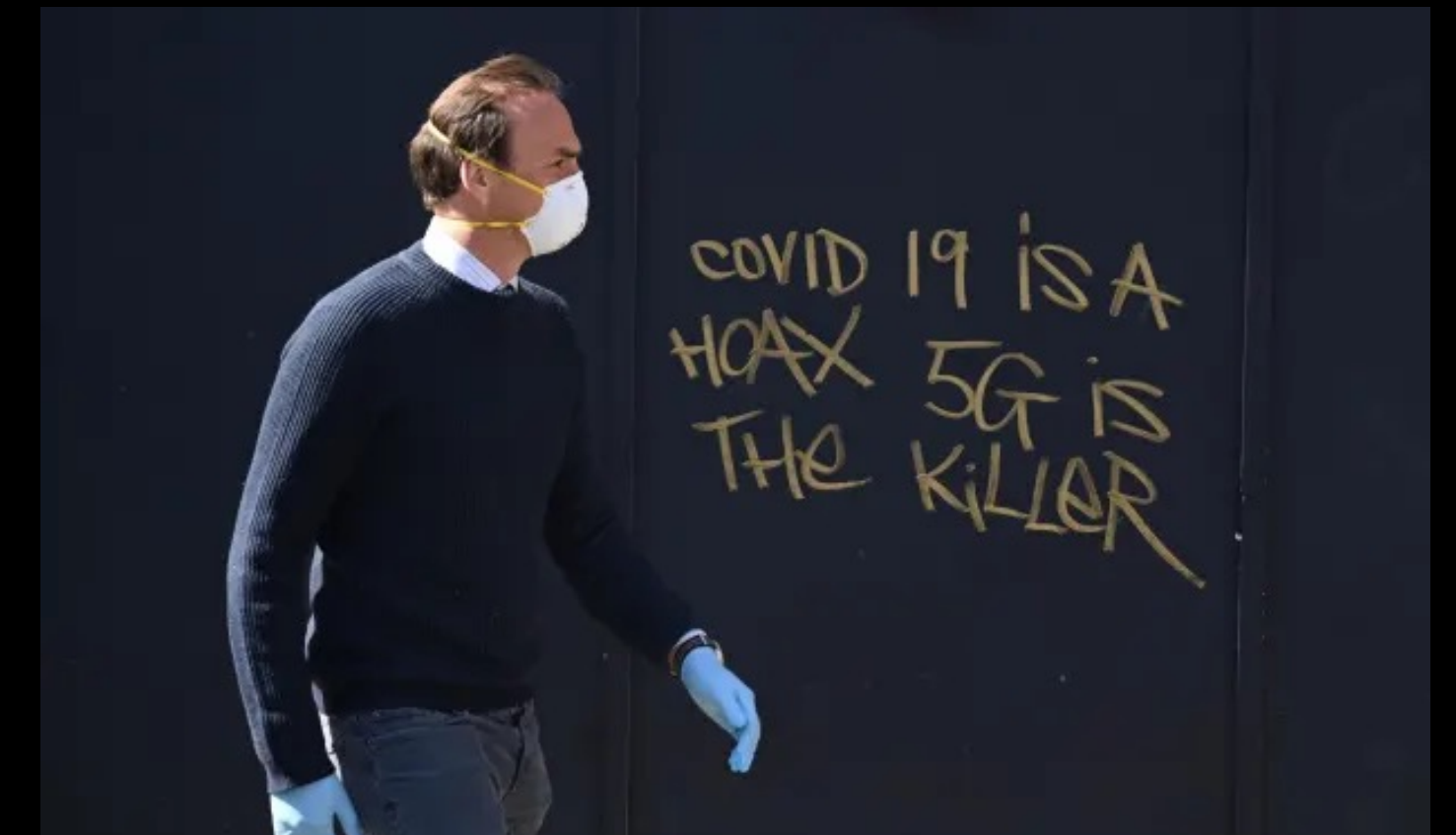


by BOB PRICE | 12 Sep 2020 | 38

LISTEN TO STORY

3:19

Big Bend Sector Border Patrol agents arrested multiple groups of migrants who illegally crossed the border on Tuesday. The groups of migrants included two previously deported men with convictions for rape and sexual assault.



Obdachlosen angezündet

Merkel machte 2015 Selfie mit einem der Täter!



Motivation: Why should we care about disinformation?

- Disinformation erodes trust in the socio-political institutions (e.g., news, scientists, experts and fellow citizens) which are the fabric of US Democracy and our society.
- Ultimately disinformation will (or has) destroy(ed) our shared reality
- Reminder: Disinformation has been around for a long time



So why is this suddenly an existential threat to society?

- More of it now?
- Seems to be made worse by technological innovations
 - Algorithmic Targeting and Personalization
 - Optimized for ad revenue and engagement
- *Disinformation increases engagement*
- Coupled with Lack of Digital Literacy
 - THIS DOES NOT EXPLAIN EVERYTHING!



The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Challenges in detecting disinformation at scale

- Massive ecosystem of disinformation to contend with (e.g., fake news, junk science, memes, Aunt Patty, self-proclaimed SME)
- Constantly Evolving and advances in AI e.g., GANs and GPT empower bad actors faster than countermeasures can keep up with
- Disinformation and detection are adversarial challenges

Promising Directions for Detection

- May leverage context/features instead of content e.g., transmission speed, number of retweets, depth of cascades
- Other potential features: network topology, origin, similarity to known messages, whether humans or bots are spreading them etc.

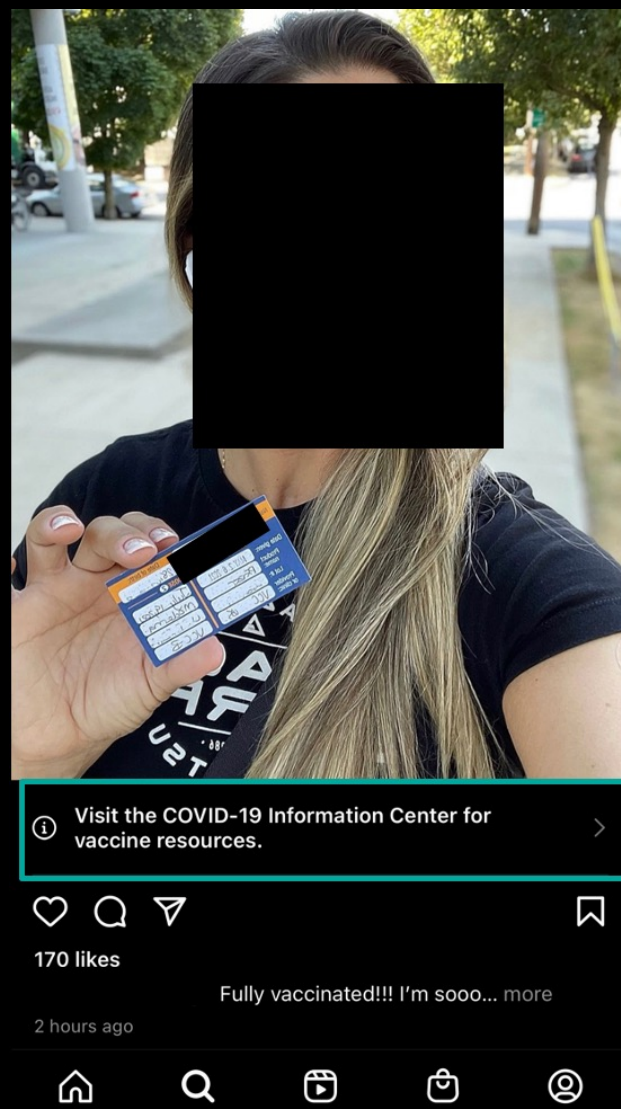
Science

The spread of true and false news online

Soroush Vosoughi, Deb Roy and Sinan Aral

Challenges in verifying disinformation at scale

- Even just fact-checking is an unsolved problem in ML. State of the art: check against known DB/Stance Detection.
- Disinformation is even more challenging to identify because AI is bad at intuition, sarcasm and subtlety. All the things that make good disinformation.



Obdachlosen angezündet
Merkel machte 2015 Selfie mit einem der Täter!



The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
- 2. Measures of Impact for Disinformation**
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Challenges in Measuring the Impact for Disinformation

- Currently no way to measure the effect or impact of disinformation on individuals or on society
- Especially quantifying second-order effects on social norms, ideologies, epistemologies, and polarization

Possible directions for metrics

- Formal statistical causal inference on human belief dynamics
- Far too many independent variables whose influences are unmeasurable



Two possible directions for metrics

- Traditional A/B testing on social media “sandboxes” but this has deep ethical concerns.

The screenshot shows the front page of a research article in the journal 'Studies in Conflict & Terrorism'. The article title is 'New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis' by Erin Saltman, Farshad Kooti, and Karly Vockery. It includes a sidebar with metrics (3,406 views, 0 CrossRef citations, 67 Altmetric), a 'Listen' button, and a bottom navigation bar with links for 'Full Article', 'Figures & data', 'Citations', 'Metrics', 'Licensing', 'Reprints & Permissions', and 'PDF | EPUB'.

Studies in Conflict & Terrorism >
Latest Articles

Enter keywords, authors, DOI, OR

Submit an article Journal homepage

3,406 Views
0 CrossRef citations to date
67 Altmetric

Listen

Research Article

New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis

Erin Saltman , Farshad Kooti & Karly Vockery

Received 10 Oct 2020, Accepted 13 Dec 2020, Published online: 30 Mar 2021

Download citation <https://doi.org/10.1080/1057610X.2021.1888404> Check for updates

Full Article Figures & data Citations Metrics Licensing Reprints & Permissions PDF | EPUB

The Agenda:

Six Strategic Targets of Support and Research

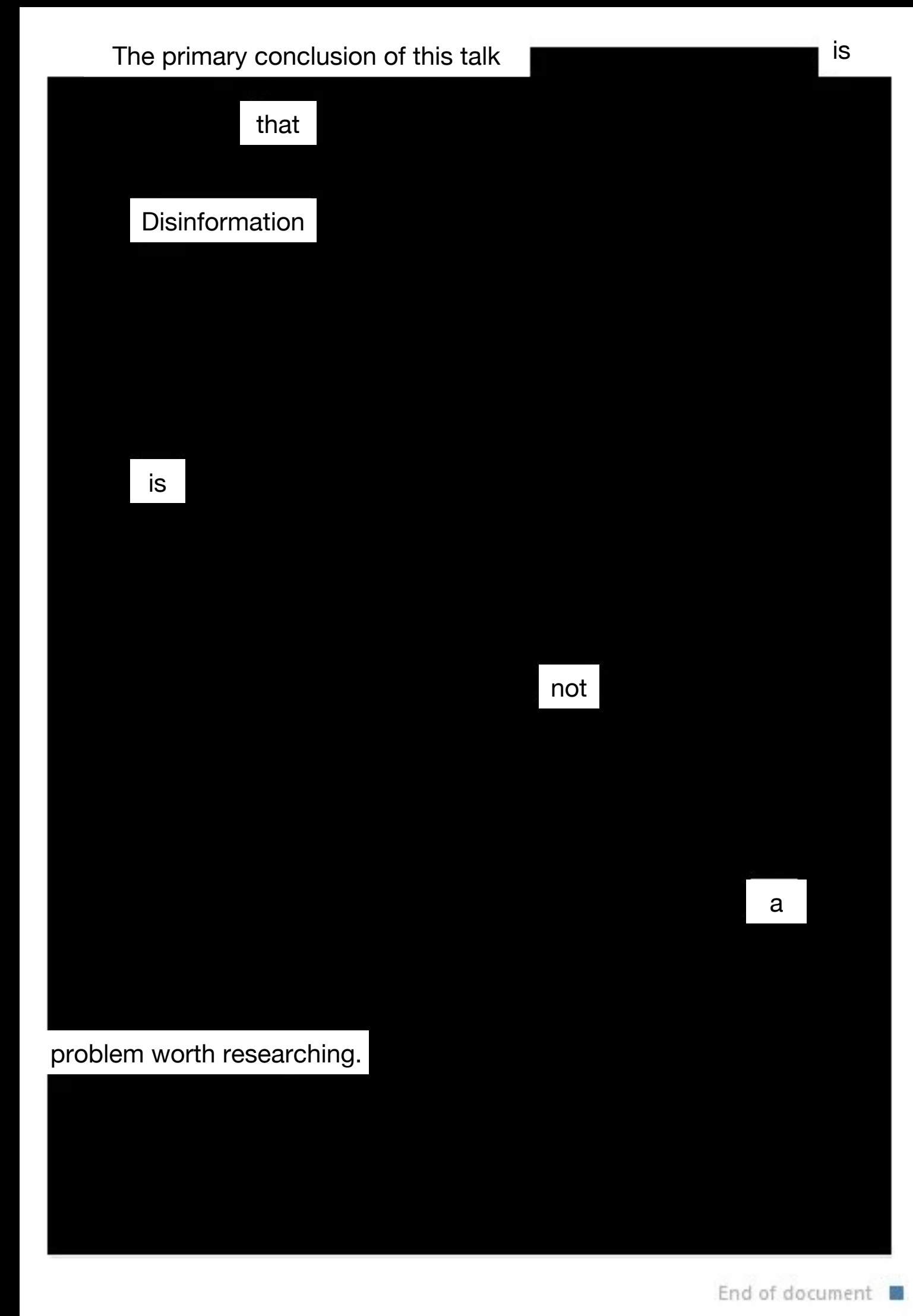
1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Current Data/Research Infrastructure

- Each group has their own data obtained in various ways (API calls, hoses, scraping, bots ...)
- Usually, the resulting data is kept private
- Makes reproducibility and replicability next to impossible
- Usually, this data represents some small fraction of the story

Current Data/Research Infrastructure

- Thought exercise: Imagine you need to understand a report, but you are only given 1% of the text.
- Makes concluding anything meaningful challenging, also further challenges related to reproducibility



Open Research (and Data) Infrastructure

- Need a common research infrastructure to access data from each platform under ethical guidelines that protect user privacy
- This could be accomplished with:
 - Comparable free open APIs for each platform with vetted access
 - Protocols and large-scale infrastructure so citizens can contribute data for research in a secure and privacy preserving manner

Challenges with Open Infrastructure

Many legal, ethical and practical challenges with such an infrastructure

- How to ensure bad actors are not allowed access?
- Who controls access and vetting protocols?
- How to maintain users privacy as well as right to be forgotten?

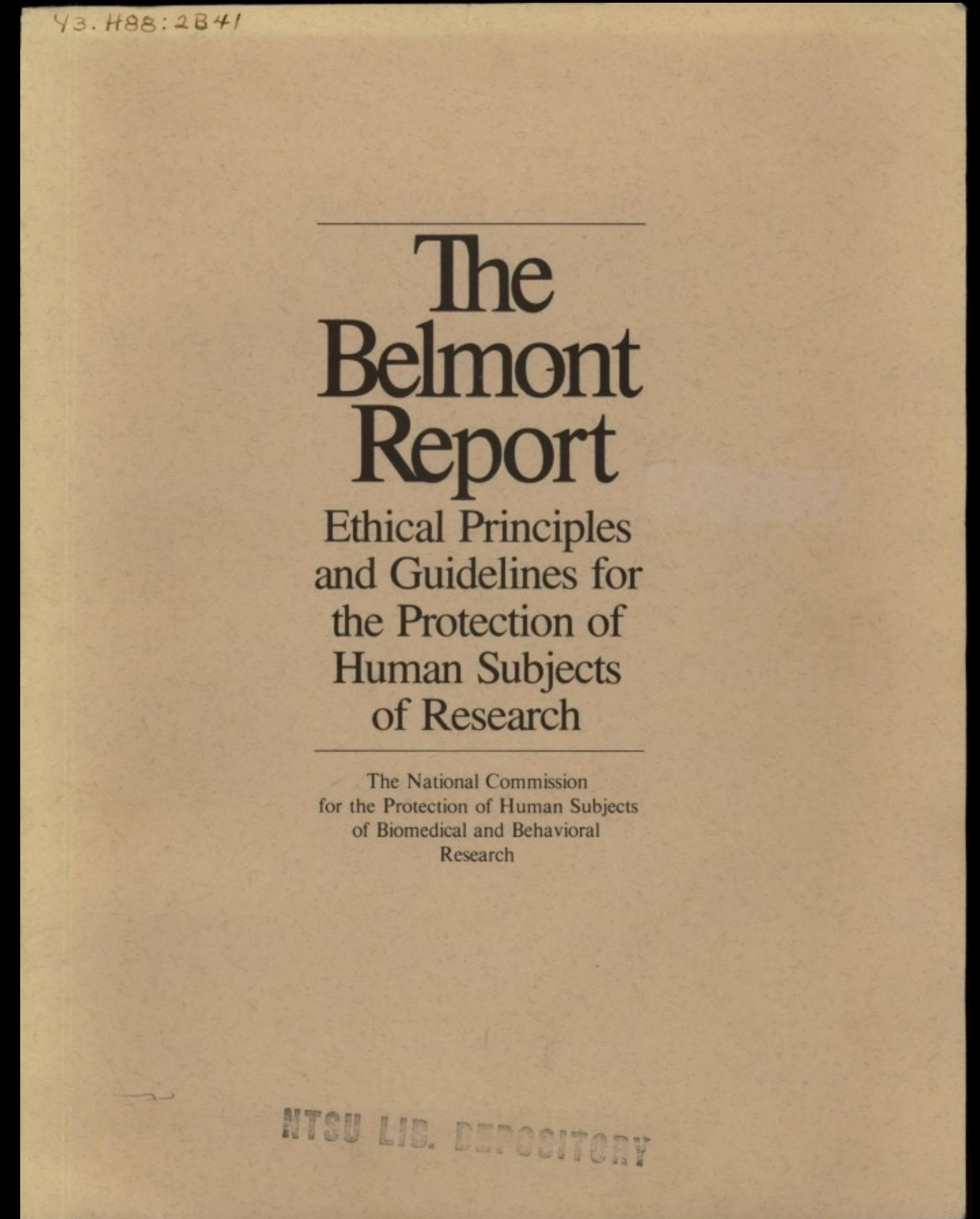
The Agenda:

Six Strategic Targets of Support and Research

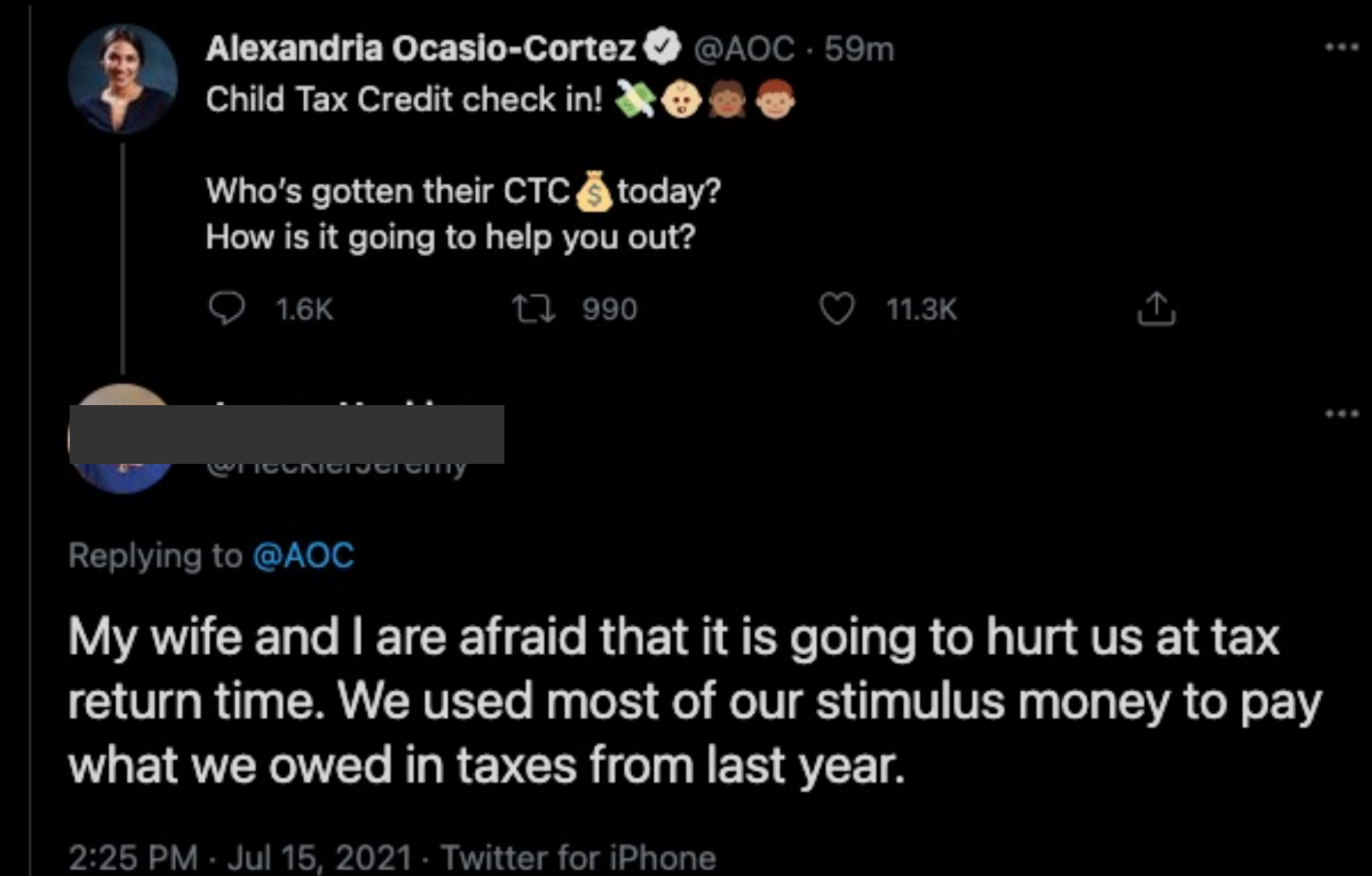
1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
- 4. New Ethical Guidelines**
5. Educational Interventions
6. Workforce Training

Current Ethical Guidelines

- 1978 Belmont Report is the current standard for Humans Subject Research.
- How this extends to the digital world is unclear and applied differently by different organizations.



How does Belmont Report apply?!



New Ethical Guidelines

- US policy makers have an opportunity to commission a report of similar impact to the Belmont report, updating how its ethical principles should be interpreted in the current digital world.
- Such guidelines would ensure transparency, fairness and minimization of harms in this necessary research agenda.



The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Educational Interventions



HOW TO SPOT FAKE NEWS

- **CONSIDER THE SOURCE**
Click away from the story to investigate the site, its mission and its contact info.
- **READ BEYOND**
Headlines can be outrageous in an effort to get clicks. What's the whole story?
- **CHECK THE AUTHOR**
Do a quick search on the author. Are they credible? Are they real?
- **SUPPORTING SOURCES?**
Click on those links. Determine if the info given actually supports the story.
- **CHECK THE DATE**
Reposting old news stories doesn't mean they're relevant to current events.
- **IS IT A JOKE?**
If it is too outlandish, it might be satire. Research the site and author to be sure.
- **CHECK YOUR BIASES**
Consider if your own beliefs could affect your judgement.
- **ASK THE EXPERTS**
Ask a librarian, or consult a fact-checking site.

The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Workforce Ethics Training

- Much of the technology that is being blamed today for disinformation manipulation were developed with benign intent.
- The technology being designed today like GAN/GPT are being used in psychological warfare but we don't think about CS students as weapons developers
- Need to train computing professionals to consider the potential misuse of their algorithms including applied ethics training



The Agenda:

Six Strategic Targets of Support and Research

1. Detection of Disinformation at Scale
2. Measures of Impact for Disinformation
3. Open Research Data Infrastructure
4. New Ethical Guidelines
5. Educational Interventions
6. Workforce Training

Thank you for your time.
Questions?

Joshua Garland
Santa Fe Institute
Joshua@santafe.edu

