CONTENT MODERATION

Ideas to Counter

Wittaker's "claimsed" consistent with Vraga & Bode's (2020) "observational correction." https://sign.ashpublications.org/ebh/481/E2052/LRH.2020.305918

Technique-based inoculation messages can be useful in forewarning the public about NORC's framing devices (anecdote, uncertainty, etc.).

global governance networks need to go beyond representation, how open are we to South Africans, Kenyans, of concepts and methods?

Provide citizen scientists with structured annotation tools empowering them to engage in open, de-biased sense-making/moderation

Support for diverse (Global South) tech whistleblowers and leakers, as well as real advocates centering the global North (this means strong security, worker advocacy, international aly ship (outside of their institution)

trust network-building with more open-ended outcomes but academics, journalists, and other people who connect and bond over access to support: legal assistance, cybersecurity, worker advocacy, international aly ship (outside of their institution)

Overall theme (Jonathan Ong): asking how people outside centers of academic power can be supported to take leadership on this issue.

Global media storytelling of tech/democracy space should go beyond dystopia disenvowoning horror stories that alienate audiences

NORC's social data collaboratory and Ameripaek panel also provides a representative samples of social data for analysis.

Adams' "Public Editor" sounds similar to "Fiskkit" https://fiskkit.com/about

A difference is that Public Editor assigns annotators randomly, requiring articles to check (so it is difficult for individuals to distort overall results for an article), while Fiskkit encourages people to pick own articles and annotate the whole article.

Is it possible to usefully provide incentives for annotation, perhaps in the social-media platforms to get *them* to experiment with content moderation techniques that optimize for such goals as (1) shared public understanding of all the participants, (2) mutual and an understanding of the points of view that a social-media user *doesn't* agree with, (3) the de-demonization of those who one disagrees with (perhaps as indicated by the use of language).

question from audience: Would it be possible to create a collaborative interactive library online that would organize information to ensure that everyone is exposed to it? Digital interactions preclude the possibility for inclusive participation and shared experiences because choices who, where, when, and so on? - Jeanne

from audience: Differential privacy seems to be an interesting framework for responsible data sharing and access. See whitepaper from Spectus.

Can we empower NGOs — perhaps backed up by random-sampled deliberative democracy work — to provide the needed oversight to all of the content moderation? Who else (or who tried)? Such oversight might be first experimented with by the EU in their efforts to make social platform audits

non-state-led networked governance structure to help facilitate the detection, analysis, and mitigation of evolving threats of disinformation (propaganda) is a worthwhile endeavor. Lessons on history of similar efforts in cybersecurity focused on narrow harms and actors' economic incentives

Obstacles and Challenges

An obstacle is in focus. Mal-Mis-Disinformation are not the same nor do they work the same or have the same consequences. Intent is different also. Need to focus effort on the most dangerous

As Matias, notes, we need to include community voices as in civic science.

funding of disinfo research captured by platforms, states, intelligence, narrows down rather than broadens solutions space.

"Content moderation" implies surveillance & policing of speech, only one among many scientifically-tested moderation options in use

Platform data access problems make research extremely difficult. Platforms door but need something like FACA in US.

2 heads are not always better than 1, let alone 2000 heads. Collective intelligence requires careful architecting to perform well and earn trust at scale

This is a good point! I think that how fine and users experience the platform matters a lot. Reddit is with true-false and that has to do with digital and methods literacy. As well, the methods to support good judgment are the best practices from the field of health communication. Indeed, Reddit's moderation is like this. This all has to power

For Brigham, what lessons have you taken from citizen journalism (blogs, wikis, etc.) for your work? Some of this leads to the risk of mis-/dis-/malinfo...

Disinformation is about power of narrative not a true-false distinction.

Platfrom data access is an overwhelming need. DSA opens doors. Need PAT.

Previous amounts of "fake news" or disinformation taken out of context, or amplified in its importance, undermine this. This is a challenge for notions of enforceable community standards on platforms.

For J. Nathan Matias, I like the idea of "sociotechnical" but the devil's in the details right? How can we bring these different groups together? Also, for "Independent research" - how can we do this when there are such strong gatekeepers to data, especially social media data?

There are some good lessons from other transnational institutions on challenges of action collectives. E.g., the Internet Routing Registry (IRR) system's role in mitigating the security vulnerabilities inherent in the Internet routing. A global socio-technical system plagued by misaligned actor incentives, collective action costs, and unmanageable interdependencies. See https://www.arpnet.com/slides/RO.NANOG72F0807.20171290292

Northern orgs focus on more legislation/regulation. But global majority has been concerned with this approach. Restrictive laws are often used against gender activists, who can be targets of disinfo or accusations of being agents. - Paula, Association for Progressive Comms.

Audience question during Brigham's talk: How with such a granular approach do you avoid fallacies of composition and division? - Paul Syverson

How do we ensure access for all communities to contribute to publicly-annotated systems?

With Public Editor, we invite literally everyone to participate. Onboarding requires some instructional and gamification before a person is accepted as an annotator. Then, we monitor annotator performance to ensure they aren't becoming a rogue strategy. Annotators get a scored experience as a function of how well they are learning to discern & label clues of disinformation for their compatriots. They badge up, and ultimately earn certificates of discernment they can post on their LinkedIn profile (for example).

Open Questions for

solutions dont always need to be scalable and transposable, healing and resilience in communities can be hyperlocal and that's ok

Moderation often opposed on free speech grounds. Why is content "amplification" less of a focus? There is no legal protection to amplifying harmful messages.

How can approaches to disinfo work arise from affected communities & their ways of seeing (Ong)

To what degree do approaches send signals about democratic norms for countering speech? How do authoritarian governments approached? How can approaches reflect strong democratic values? Jonathan notes these

how to develop solutions/interventions in tech spaces that are openly hostile to anti-disinfo work? -Curt Lindley

Brig: we collaborate with citizen journalism orgs like the UK's Bylines Network and are in talks with Wikimedia Foundation about how these tools can be integrated. Such Wiki collaborations have a high ceiling of legitimacy, because they are open and inclusive. But we really need to 'nail' the architecting to ensure we are creating 'collective intelligence' and 'crowd wisdom' and not the 'madness of crowds' plays into the current narrative wars.' Happy to follow up in-depth on this in a medium that allows for it.

1. The annotation schema should evolve with disinformation trends. 2. Annotator biases should be monitored & mitigated. 3. Output should be used to

Yes, the labeling process is 1 place where bias is introduced. Agree!

Amelia Burke-Garcia - I like your message framing approach! I wonder how you detect anomalies and adherences to overall annotations. Great questions! "Evidence" module guides annotators to distinguish "anecdote not data! ... and our "Quoted Source" module guides annotators to identify, label, and assess the expertise/standing of quoted people, while the rest of the system addresses the quality of their statements. These assessments are separated from the assessment of an article, so that journalists are rewarded if they express appropriate skepticism to an idea from a misinforming source.

We used a set of labelers to label samples of social media data, then ran inter-rater reliability scores. The frames and definitions of the frames were pulled from sources like the US Surgeon General's handbook, amongst others.

We know "jaebonding" is an enormous problem in countries like the US, where people live in an ideal sense, how do we want gov't-platform communication to take place, through what channels, with what degree of transparency?

Absolutely. This is why fact-checking is not enough. 'True facts' are often manshaved and placed in a misleading context... and Public Editor's 'reasoning-process' is explicitly designed to address such issues... (with respect and admiration for fact-checkers with whom we collaborate.)

Perhaps the tie of these systems forward is a transparency approach to this problem, so that students in every community are able to participate (and to learn about the elements of the needed critical thinking). Brigham also just echoed Dr. Cooke's suggestion of libraries.

With respect to true information out there, Wikipedia's "fact checking" tests of an article is whether it properly provides context for the claim. Perhaps we should be asking how important is this article's point relative to the large context that it is part of. And has this been made clear in the article or is this being distorted.

Our team has begun to converge on this message: We are sometimes misled by false or manipulated evidence. But we are more often mislead by omissions, contexts and ways of framing (selection of evidence) due to distorted or manipulated frames. Disinformation works both by manipulating evidence and by manipulating frames. We need to focus on the manipulation of frames from the manipulations of facts.

Other

Re: Wittaker's "claimsed." Why limit to Twitter? Reddit? https://twitter.com/philmcintosh/status/1270000000000000000 and https://www.wired.com/2019/04/1TOh-mize-twitter-account-hype-science-reporting/

It will be available on NORC's website, using an interface. We also hope to make the source code available

Learn more about such a system at https://publiceditor.io

There, individuals and classrooms/orgs can onboard into the system to contribute

Curious how NORC's model will be applied: is it a browser plug-in?

Wondering about the choices for posture/voice for agents – shouldn't they be tailored to the person engaging with the agent?

Resource suggestion from audience: documentary "The Antisocial Network. Memes to Mayhem" to better understand disinformation in social media.

Can suppression of national events by media outlets, like recently with the protests in Brazil, be considered misinformation by omission? - Solomon

u

EDUCATIONAL INTERVENTIONS

Ideas to Counter

Awareness of actors and tactics

- Teen focus -Multidisciplinary approaches -Project based

Incorporate contemporary pedagogies into classroom instruction

Teach discipline-specific credibility criteria (e.g., in science, focused on peer review & balances) in classes

Identify protective thinking & motivated reasoning still shape source evaluation & info eval.

Greenspan & Biblehoover: What evidence do you have that it is effective and what do you do good habits?

Teaching students about mechanisms of information flows: on social media, algorithms, echo chambers, etc.)

Inoculation - warn individuals about misinformation they expect to infiltrate, sources that might be misleading

NORC's framing devices (anecdote, etc.) could be utilized in technique-based inoculation messages.

Before developing a real or virtual experience, or demonstration method and with Scientific Information), ask to the students in groups of 3-4 students to develop prepared questionnaire, which is their level of knowledge of the subject matter, the theme of the experiments. Ask for the same questions at the end of the experiments. Comparing Compare. Example: is the Earth flat or (almost) spherical ?

We build resilience when we empower social media users. - Rachele from Truth Bomb Social

Obstacles and Challenges

Scale: Are the benefits worth the cost?

Already crowded curriculum / standards

Hard to fully divorce emotions and values from politics (e.g., via political implications of climate change acting as activators, etc.) - Schaufele (2014)

Experiences that foster virtues like intellectual humility, curiosity

Pearcentre: The examples seem special. What general principles are we seeking to develop?

Grob's (post)-media literacy: Are the techniques are evolving so quickly, how can we hope to keep up with the advancements?

More challenging than identifying fake text? Is there a hierarchy of thing to look for?

- Matt Groh

We can keep up with advances by putting attention towards these advances, keeping track of what matters and communicating in an effective way. Not everyone needs to keep up with every details, but there can be an organization paying attention and sharing resources with teachers - Matt Groh

I really appreciated that Verich & Blackburn's piece on lateral reading led extraordinarily. This status would seem to position it somewhere. Lateral reading is one of the basic culture war battles over curricula and books in classrooms and therefore that are becoming contested across the U.S.

In an online environment full of bots that incentivize the creation and spread of disinformation, to what extent is consumer media literacy a individual solution to a structural problem?

Grob

I know dwash boyd has critiqued traditional media literacy approaches because of production skills as having the potential to backfire if not done well — she gives the example of young people who participate in protesting police violence and redress using the knowledge they gained. To put it bluntly, how do you set up a formal or informal curriculum that motivates people to use their knowledge of generative AI for good and not evil?

boyd, danah. "You Think You Want Media Literacy… Do You?" Data & Society Blog Posts, March 16, 2018. https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af78b67

Open Questions for

Does awareness catalyze behavior change?

What are the best approaches for scaling awareness?

How many "covered" does it take to minimize the negative impacts of

What is the optimal age or audience for awareness interventions?

It depends on the context and the goal. It is hard to distinguish true fake video for domains like political speeches. https://mathgrih.runopdb.net/viral_videos/fake_political_speeches.pdf

- Matt Groh

Is media literacy education more effective as a stand-alone course or integrated into discipline-specific courses?

What are the differences in the effects of educational interventions when it comes to individuals evaluating information that aligns / doesn't align with their existing worldview / political ideologies?

Do educational interventions about misinformation change students' information seeking / evaluating behavior outside the educational context and in everyday internet use?

Impacts of educational interventions when evaluating different types of media (e.g., video, text, image)?

To train people (teachers, journalists, and the general public) how to obtain information to counteract misinformation, especially the example of Google produced in-app. Fact Check Tools

Is it necessary to divorce media/digital scientific literacy from consideration of real-world impacting readers in the real world?

Sam Wineburg: What are the most important things for people to know so we can just have time to work on a few things?

Is it helpful to provide tools that can be taught in a classroom setting and applied consequentially outside the classroom? Do students learn worse if their exercises have real world impact?

Should digital literacy be taught in real-time? By analogy: Google Maps does not teach people to memorize a map of their city. It gives them turn-by-turn directions to navigate their world. Presumably, Google Maps does this in a way that works effectively. But does it atrophy navigation skills? Perhaps it delays their growth but ultimately supports them(?). What about in the information space? Should people learn to do lateral reading/verification in a sterile classroom and then apply it in the world without help? Or, should they get some kind of real-time signals to navigate the 'biasphere' as they go.

Would it be possible to do educational research to find the right balances between 1) teaching how science works at its best to reach better understanding and 2) teaching all of the appropriate skepticism of much online information?

Other

How to connect these educational efforts (The Diaintg Project, UW's MisinfoDay, etc.) to new programs trying to launch? Can NAS be a clearinghouse for resources?

How can multi-sectoral organizations leverage the TDP network as a feedback mechanism to test these hypothesis and gain feedback on what works and for whether certain materials resonate with their demographic?

From audience: Global Engagement Center from the U.S. Department of State is a good educational resource.

resources from audience: I've found Ignorance of The Constitution and amendments contribute to many of the comment on social media. Two articles underline the problem:
"Only 26 percent of Americans know the First Amendment" by Ariana Pekary, Pew Research Center, 2019
Civic education must be comprehensive and engaging to prepare the next generation to be knowledgeable and active leaders.
"Civic Education Research Review" https://www.iacsc.org/research-review/
"Civic knowledge and public engagement is at an all-time low. A 2016 survey by the Annenberg Public Policy Center found that only 26 percent of Americans can name the three branches of government, which was a significant decline from previous years."
Cited by
"Democratic education for service members to close the civil-military divide By Aryn McBreen Spector in Stars and Stripes + July 19, 2023 https://www.stripes.com/opinion/2023-07-19/stripes-view-the-civil-military-divide-3175854.html
"Our foundational step toward this goal is guaranteeing that they fully understand the constitutional rights and responsibilities of citizenship and civics. General civic knowledge correlates with higher political awareness and participation as well as a higher reported level of patriotism. The military should be a leader in civic education and should see it as a wasted opportunity to underscore the vital importance of having not only a good soldier, but also a good citizen."

In experiments and research into teaching "scientific thinking" (see nobelprize.org/scientific-thinking-for-life/ and the work of Carl Wieman's sensibility.berkeley.edu at the university level) we have been finding it necessary to simultaneously teach some of the can-do capabilities of science at the same time as we teach all of the skepticisms — otherwise we find the students saying "I know I can't trust any claims and that "your guess is as good as mine"

Public Editor would love for any/all of these literacy/education projects to use/critique/improve PE in its crowdsourcing efforts. Our mission — to discern, & help their compatriots do so... impacting timely, consequential content that affects people's decision-making and lives (from public health to policy choices)

TECHNOLOGICAL INTERVENTIONS

Ideas to Counter

Some state legislation is being developed to support and protect authenticity marking

Idea behind the Coalition for Content Provenance and Authenticity (C2PA)

Industry-Academia collaborations to incorporate transparency labels on the internet for consumers re: advertising on misinformation

Create publicly available "native advertising" database.

Develop machine-learning approach to detect "native advertising" disclosure labels on social media.

Integrate discourse theory, persuation theory, and other communication theories into human-guided, science-based ML techniques for robust misinformation detection

DISARM Foundation provides an open framework of disinformation indicators and tactics... and counter-tactics to disrupt the supply side of intentional disinformers – https://www.disarm.foun

Make the flow of information and persuasion campaigns visible to the public. This might help with the longer term disinformation campaigns, which are more subtle. ("Frames")

Require searchable ad libraries

How can we use subnets, like .med or .health, to ensure the public has access to credible health information online, including LOCAL resources?

Delay social media response times by a few seconds

Only need to get ~60% of population to share an understanding of truth vs disinformation, and then get a "herd immunity" effect.

Develop a separate part of internet for public good.

Obstacles and Challenges

Collaborations with ad platforms for large-scale field experimentation

Lack of standardization of required disclosures.

From audience: One of the troubles with "stamping" everything with provenance info is that it can lead to very microscopic tracking of digital objects. This could potentially compromise privacy of individuals who are whistleblowers (perhaps) and everyday individuals not looking to spread disinfo. It could have a chilling effect. - Nitin Verma

For a machine-learning labeling of disinformation versus trustworthy websites, how can you provide transparency for the algorithm/training so that all (more) parties will consider the labeling legitimate?

Funding for all of these approaches? NSF divisions are not set up for this cross-disciplinary work.

This is an important issue. Best will be (at a technical level, make the neural network used open source, and (b) to show the training data. Best will be (at a particular website as legitimate or fake with a particular probability.

Everything open will be exploited.

And a complication to this is the well-known reproducibility "crisis" in computer science, where different groups using the same language model get different results.

People may ignore warnings and labels

But see evidence that pre-bunking may be effective

Don't assume a lot of effort available from the users (who are multitasking...). - "Meet the users where they are."

from audience: ongoing limitations regarding Academic development models through 3rd party outside source claiming credits long-term, allocated grants awarded yet displaced receiving end, leaving empty-handed pattern (despite?) hard earned commitments/contributions/value to disinformation and fot/ 4th-fifth Amendment violating Law Enforcement Law students during internship/ in-school assignments

Training misinformation systems is very labor intensive. (Even difficult to get enough undergrads.)

Yes, and Public Editor is doing it despite the difficulties using STOA / notation-assembly-line technology. :)

Then difficult to get sufficient uptake of, e.g., a browser extension

Yes, many people are committed to being 2nd-order rational in narrative wars. But many people are also now tired of that. If 15 - 30% of the population takes an interest in labeled content, it can tilt public discourse. The percentage points toward reason and circumspection.

from audience: Getting the most granular units-of-analysis: words in documents, and very specific misinforming phrases, is a hard problem. Stemming, (confusing correlation and causation?). Then, can aggregate up to article-credibility, journalist credibility, news source credibility. Public Editor does this.

Propose that it is best to use the most granular units-of-analysis: words in documents, and very specific misforming phrases, is a hard problem. Stemming, (confusing correlation and causation?). Then, can aggregate up to article-credibility, journalist credibility, news source credibility. Public Editor does this.

Open Questions for

Do corrections to mis/disinformation by themselves become 'cues' and misinforming by people? (Ex. Inoculating messages itself could become 'misinformation' for some people.)
"Clorox will Not protect from Covid Infection" → "for some, fail to use negation and recall the corrective information as 'misinformation' → Clorox will prevent covid infection.

Follow-up studies of effect of such EU audits

Brands have been saying they want greater transparency in ad placement — and adtech firms have been saying they would provide it — for the better part of a decade now. And there's been a lot of sustained activism around the issue by folks like Nandini in that time. Why these issues are so resistant to reform?

What is the "pick up" of the different techniques, and do some levels of the info ecosystem get better uptake than others?

Public Editor is collaborating with Wendy Hui Kong Chun (of SFU) to incorporate "authenticity appeals" into its dysinformation labeling ontology. PE is happy to collobarate with more researchers to test other communications and discourse theories for

Extremely high accuracy of labels seems to be vital. Was it true for Wikipedia? Wittgenstein(ian) who theorized that all communication would fail if the proportion of mistruth within a population's discourse grew to 50% (+/- 20%)? If labels are only 90% accurate, is that (or what) could undermine the entire

Explicability of the label-generating process also seems incredibly important. Many disinformation claims are delivered from a black box' endorsed by experts, the labels themselves constitute 'appeals to authority' and a kind of 'argument from ignorance', undermining the efficacy of

Furthermore, even if 'black box' AI-generated disinformation labels were highly accurate, what does such a scenario mean for human agency? More to the point, in a hypothetical scenario, what becomes the purpose of human agency/freedom/life if we become obsolete with respect to solving our own problems?

Other

Is the EU checking for proper labeling of "native advertising" in their audits of the platforms?

Perhaps the EU audits could be a testbed for all of the ideas presented?

Collaboration suggestion from Rachele Rice (audience, Truth Bomb Social): I lead a new nonprofit tech startup innovating to empower social media users against disinformation. Jigsaw's work and the paper I read on Prebunking was formative when we were getting started. We seek collaboration with experts and would like to connect.

Important that the accuracy of labels of disinformation be extremely high (and transparency). Otherwise, just builds distrust in everything.

From audience (Paul Syverson): Classic usable security maxim: Always design the interface so the user makes the right decision. Warning from Jeanette Wing: User error is not a diagnosis, it's a symptom.

Labels/ratings of journalists and websites can be aggregated up from disinformation labels/scores on individual articles. Journalists and articles, Research could likely show that aggregating up from specific labels at the most granular (word) level are better understood, more useful, and have higher earned

In Asia (Japan, e.g.) can require the browsers to make labels/tags visible. So solutions may be different in different parts of the world. [See also the EU DSA etc.]

A call for more funding for research in this area.

Paraphrasing Susan: In the past we have solved such problems by learning how to hold people responsible.

LEGAL AND OTHER INCENTIVES AND DISINCENTIVES FOR BEHAVIOR CHANGE

Ideas to Counter

The basic problem is that actors on the internet are not held responsible for their action. Technological fixes trying to address this, such as the proposal to open society, but creating means of holding person's responsibility for their actions online anonymity is rare in society is almost all human societies. These technological solutions do not address the basic issue.

Paraphrasing Josh Braun approaches: —"Ad Hoc" —Add friction to tight systems —Reduce complexity —Organizational culture and management techniques —Compartmentalize

From audience (Rachele Rice, Truth Bomb Social): To David Bray – inevitably the public directly is the focus of our new nonprofit tech startup. We want to help people to see disinformation and to empower them against it. One aspect to our solution is bridging experts directly to the user, and integrating AI to use the power of this technology relationally in conversations ongoing with each of our users.

Under-appreciated subsections 5230b,3&4. These sub-sections seem to offer a doorway for some of the solutions on offer in this workshop. "It is the policy of the US to encourage... technologies which maximize user control of what information is received..." and "remove disincentives for ... blocking and filtering software" I'd we can argue that user-opt-in labeling technologies are a more modest intervention than 'filtering'

Require licensing for those using "broadcast" technologies.

Paraphrasing Nathalie Smuha: EU and DSA focusses on increasing transparency of platform providers: how do they work?
"Platforms reach >10% of population, they need to do risk assessments for their platform design and algorithms — harms to human rights (including free speech) and mitigations measures. Public feedback on these assessments."
"Trusted flaggers.. Platforms have to make data available for researchers to verify what they claim they are doing. If follow the law, they are indemnified from further responsibility.

Instead of limiting or censoring free speech, we can label all the dis-/mis-information, converting it into learning opportunities for the public

DSA provides data and funding for researching 'social harms; experimenting with interventions, and even compelling very large platforms to adopt interventions proven to mitigate social harms

Develop NGO and Random-sample Citizen Assembly oversight of platform regulations. Can this be tested with the EU example, by setting up oversight of their audit's metrics of harm.

Paraphrasing David Bray: Give individuals the control of their data's use by others (and do so collectively to reduce to workload/burden)

Citizen Juries

A national registration of data brokers

230 envisions "user technologies: helping provide alternative (filtered) feeds.

Obstacles and Challenges

Section 230 of the Communications Decency act prevents holding platform responsible. That needs to be hanged and this is NOT fundamentally a free speech issue. Big Tech is not responsible or responsible for what they put out here or manufacturers for the harm their products cause. Immunity is no longer necessary for its maturity to grown.

From audience: What is truth? If a government is saying that something is fake news (but actually are just trying to hide a story), and the systems you propose take their word for it, then your attempts to keep people safe in fact did the opposite. And your credibility would be lessened, decreasing trust.

Exactly! Important to provide specific, granular labels pointing out particular misinforming frames, not simple labeling an article or news source as 'fake'

Paraphrasing Brandie Nonnecke: Platforms are likely to assess their own risks so that they can mitigate them easily.
So need more robust mechanisms for 3rd parties to identify risks.

Danger of using government regulation as a source of power.

Could this not be a "both/and"? §230 and DSA make it possible for governments to empower/encourage people-powered 3rd party (academic/non-profit) tools controlled by users as extensions to platforms.

Open Questions for

Research ideas for non-profits/labs:

how can we use GenAI to improve GenAI+human behaviors & intents identification

what are effective ISPs+Platform protocols limiting abusive GenAI movement

field and test random citizen juries included in oversight mechanisms

Early research that GenAI might help people work themselves out of conspiracy theory thinking - there is possible "hope" with GenAI too: see **Durably reducing conspiracy beliefs through dialogues with AI** https://osf.io/preprints/psyarxiv/xcwdn

Can we get better metrics on the prevalence of scams?

Explore, test, and field experiments - for example: if we allow people to join data cooperatives and chose when, where, and with whom they share their data - one could imagine our members could be beacons that send out a "do not record without my consent" signal.

If people chose to do that beacon in an airport, that might mean they have to do a special screening at the airport vs. a faster one.

Information Policies, Knowledge Overload, Limited Attention Spans, and Our Responsibilities as Information Systems Professionals https://papers.ssrn.com/sol3/papers.cfm?abstract_id=962722 - ...the average knowledge worker - someone who is part of the growing information economy - loses 2.1 hours a day to interruptions associated with multi-tasking. If those workers make an average of $21 an hour, that adds up to $588 billion a year, more than the gross domestic product of Canada. A recent study finds knowledge workers experienced interruptions approximately once every 10 minutes and it took an average of 23 minutes for them to return to their original task.

What can we do as Information Systems professionals to address Information pollution?

Research what role can the Professions and Professional Societies play.

Ethics and the Modern Professions: Autonomy, Social Institutions, and Potential Futures https://papers.ssrn.com/sol3/papers.cfm?abstract_id=884601 - Within modern society, we frequently rely on strangers for professional services. This course will examine the role of professionals and the professions in society. Specifically, this course will examine the role of ethics in maintaining autonomy of the professions as social institutions, as well as helping to ensure that we (as customers who rely on professional services) generally can place trust in professionals to do the right thing even when no one is watching - because often no one is.

Other

Is it possible to regulate amplification of speech (drowning out others' speech) without regulating speech.

Renee DiResta wrote on this "right to reach"https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/

@Jeff: Does this violate 1st Amendment? It would be like saying you can publish a book, but only sell it to four people. Of course there are time, place, and manner restrictions but I don't see this playing out in online environments (unless it's harmful content such as pornography, etc.)

A possible argument on this thread: Platforms have a different role as a public communication channel which allow them to be immune from liability, so their amplification doesn't seem to be the usual example of free speech. If they are open to liability for what they amplify then the free speech point would seem to apply.

Tackling GenAI and Bot Enabled Disinformation and Inauthentic Content

Why We Need New Forms of Deterence:
Stresses and Social Changes

Will need to democratize tradecraft of "information discernment" + whole-of-society "digital dignity" consistent with Constitution

Different Deterrence Regimes:
Whole-of-Society Solutions

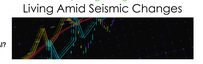
May need to launch organizations to record and show "mosaics" of designated real-world events relative to GenAI, to learn of risks, simulate scenarios, and test new tools

Will need operational NGOs to tackle issues plus more ways for individuals to have "voice" – cannot be done by gov't alone

Different Deterrence Regimes:
Abuses Counter to Civil Norms

Ideas: must avoid heavy approach that forces AI underground akin to "ham radio licenses", licenses to unleash a GenAI? certified ethical data scientist / ethical GenAI developer?

Opening Thoughts:
Living Amid Seismic Changes



Last 50 years: we've democratized tech previously just available to intelligence components of major nations in the 1970s

Different Deterrence Regimes:
Addressing GenAI Abuses

Within a Society, Gen AI abuses counter to:
... Civil Norms
... Laws
... National Defense

By:

Closing Thoughts:
More Operational Spaces

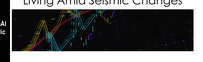
Need places public can trust to operationalize non-politically, perhaps Labs & Non-Profits with bipartisan oversight?

Different Deterrence Regimes:
Whole-of-Society Solutions

Ideas: must have national privacy protections updated for GenAI need data stakeholderism, perhaps data coops, for public potentially crowdsource spotting/tips re: GenAI abuse

By:

Opening Thoughts:
Living Amid Seismic Changes



However: we have not upgraded how to do the work of defense, security, and civil society given super-empowered populations

Premise: Countering Disinformation Requires Tackling Bots/GenAI Propagating Inauthentic Yet Human-Like Content

Free and Open Societies Cannot Tell People What to Think and Governments Cannot "Moderate Content" Lest It Look Like Censorship

Must Operationalize Ways to Galvanize Either Individual "Information Discernment" or Whole-of-Society Community Actions to Tackle Bots/GenAI Propagating Inauthentic Content

Why We Need New Forms of Deterence:
Stresses and Social Changes

Bad actors will use tech to attack U.S. national strengths in ways we cannot similarly use tech to counter → need new strategies

Noting there is possible "hope" with GenAI too: see **Durably reducing conspiracy beliefs through dialogues with AI** https://osf.io/preprints/psyarxiv/xcwdn

PROPAGANDA For High School Kids In 1948! Would They Show This Today? https://www.youtube.com/watch?v=8AGpILvdwDM