Icebreaker

Where do you or your partners and target groups get news?

https://padlet.com/swapneelsmehta/where-do-you-and-your-partners-target-groups-get-your-news-f-s2lo2ggbkcuj57hx

Disinformation: Threats and Mitigation

Friday 24/03/2023   Swapneel Mehta   One Fact Foundation

"Challenges and Opportunities for Media Development"

NEWS

FAKE

About Me

2019-23	Data Science Ph.D. Candidate @ New York University
2022	Early Detection of Misinfo. @ Twitter Civic Integrity
2021	Trending Hashtag Recommendation @ Adobe
2018-19	Machine Learning x Particle Physics @ CERN

@swapneel_mehta

Mis, Dis and Malinformation

MDM Archetypes

Misinformation

Inaccurate or misleading
information (no intent)

Disinformation

Misinformation +
intent to deceive

Malinformation

True information +
intent to deceive

Mis, Dis and Malinformation

MDM Archetypes

Misinformation ▪

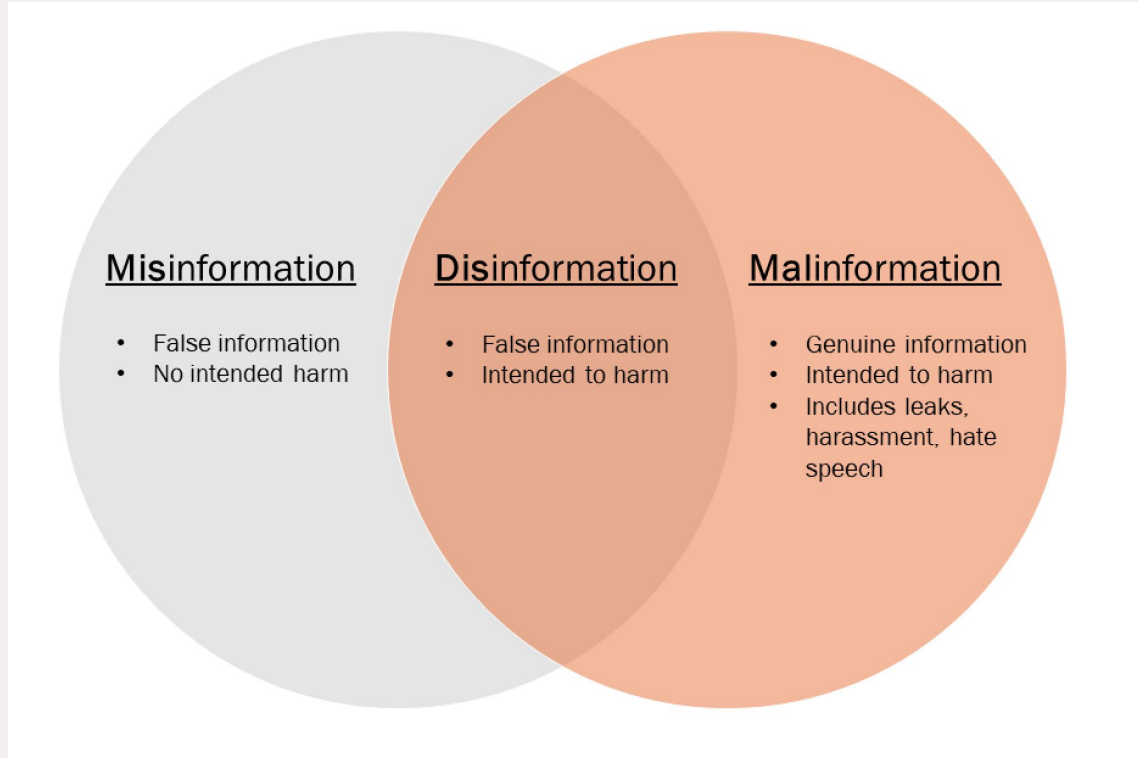Inaccurate or misleading information (no intent)

Disinformation

Misinformation + intent to deceive

Malinformation ▪

True information + intent to deceive

Mis, Dis and Malinformation

MDM Archetypes

Source: Council of Europe's Information Disorder framework.

What would you classify this as?

NEWSMAX   TV   HEALTH   FINANCE   WORLD

NEWSMAX

Monday, March 20, 2023

Home | Platinum | Newsfront | Politics | Opinion | Podcast | The Wire | Books | Best Lists | Specials | Sci & Tech

Home | Newsfront

Tags: svb | black lives matter | esg | social justice

SVB Donated $73M to 'BLM Movement'

Free Newsmax E-Alerts

Email:
Country: United States
Zip Code:
SIGN UP

Privacy: We never share your email.

https://app.sli.do/event/1hXNLSYVS2d9Uy8vGeuB7N/embed/polls/d14a6f68-6a06-49b9-9662-a4b07d047215

What would you classify this as?

MDM Examples

What would you classify this as?

MDM Examples

Donald J. Trump ✓
@realDonaldTrump

Heartwarming to see all of the tremendous support out there, especially the organic Rallies that are springing up all over the Country, including a big one on Saturday in D.C. I may even try to stop by and say hello. This Election was Rigged, from Dominion all the way up & down!

⚠ This claim about election fraud is disputed

What would you classify this as?

Let's watch the Video

https://app.sli.do/event/1hXNLSYVS2d9Uy8vGeuB7N/embed/polls/49e7e3e5-0de6-4841-ba9a-e90c019a5252

What would you classify this as?

MDM Examples

Let's watch the Video

Modalities of MDM Operations

Text

Audio

Images

Video

Thinking through a Historical Lens

Text   Newspapers

Image   Paintings

Audio   Radio

Video   Television

Modalities of Disinformation

1999-07 Texting

2003-08 Image-sharing

2000-12 Podcasts

2005-10 Video

Discussion

Which modality is most effective to spread disinformation?

Discussion

Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps?

S Shyam Sundar ✉, Maria D Molina, Eugene Cho

Takeaway

- **Video is more <u>persuasive</u> than text and audio:**

  - Text needs people to read, visualize, imagine, and understand
  - Video is upfront, 'see to believe'
  - Video is associated with greater engagement and higher perceived realism

Discussion

Why do people share fake news?

Discussion

The psychological drivers of misinformation belief and its resistance to correction

Ullrich K. H. Ecker ✉, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga & Michelle A. Amazeen

Takeaway

Cognitive drivers

Intuitive thinking
- Lack of analytical thinking and/or deliberation

Cognitive failures
- Neglect source cues and/or knowledge
- Forget source and/or counter-evidence

Illusory truth
- Familiarity
- Fluency
- Cohesion

False beliefs

Socio-affective drivers

Source cues
- Elites
- In-group
- Attractive

Emotion
- Emotive information
- Emotional state

Worldview
- Personal views
- Partisanship

Methods of Proliferation of MDM Ops

Adversarial Actors	
Agenda	**Approach**
Dismiss	Discrediting true information that contradicts disinfo.
Distort	Taking true info. and altering or exaggerating it.
Distract	Deliberately lying / creating false info. to deceive others.
Dismay	Using scare tactics to prevent the truth from coming out
Divide	Employ polarization tactics to exploit divisive opinions

Ben Nimmo, Graphika

Manifestation of MDM Ops

Adversarial Behaviors	
Behavior	**Description**
SockPuppet	Online identity used to deceive, typically in public
Astroturfing	Taking true info. and altering or exaggerating it.
Sealioning	Pester people with persistent requests for evidence (JAQ)
Catfishing	Online identity used to deceive, targeting individuals

UNHCR,

Manifestation of MDM Ops

Table 1 Communication objectives BEND

		Manipulating the narrative		Manipulating the social network
Positive	**Engage**	Messages that bring up a related but relevant topic	**Back**	Actions that increase the importance of the opinion leader or create a new opinion leader
	Explain	Messages that provides details on or elaborate the topic	**Build**	Actions that create a group or the appearance of a group
	Excite	messages that elicit a positive emotion such as joy or excitement	**Bridge**	Actions that build a connection between two or more groups
	Enhance	Messages that encourage the topic-group to continue with the topic	**Boost**	Actions that grow the size of the group or make it appear that it has grown
Negative	**Dismiss**	Messages about why the topic is not important	**Neutralize**	Actions decrease the importance of the opinion leader
	Distort	Messages that alter the main message of the topic	**Nuke**	Actions that lead to a group being dismantled or breaking up, or appearing to be broken up
	Dismay	Messages that elicit a negative emotion such as sadness or anger	**Narrow**	Actions that lead to a group becoming sequestered from other groups or marginalized
	Distract	Discussion about a totally different topic and irrelevant	**Neglect**	Actions that reduce the size of the group or make it appear that the group has grown smaller

Social Cybersecurity

Countering MDM Operations

So how can we stop these tactics?

Discussion

-

Shifting attention to accuracy can reduce misinformation online

Gordon Pennycook ✉, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles & David G. Rand ✉

Takeaway

- **People say sharing only accurate information is important**

    - Sharing does not necessarily reflect belief
    - Unique intervention — follow people and prompt them to identify accuracy of headline
    - Improves the quality of shared information after the intervention
    - Do people just need nudges?

Identification and Mitigation

Category	Approach	Examples
Fact-checking	Manually verify claims in posts	IFCN, Open Feedback, Politifact, Snopes, FactCheck.org,
Scores	Scoring the credibility of media providers or accounts	MBFC, Iffy News, NewsGuard, Botometer, Have I shared Fake News
Policy Interventions	Platform content moderation, external	Twitter, Meta, Reddit, Media Literacy (Interventions)
External datasets and dashboard audits	Build bespoke games, dashboards for civic datasets	Bellingcat Ukraine, Hoaxy, TBCOV, Bad News Game, Inoculation Game
Coordinated Inauthentic Behavior	Detecting Networks of superspreaders and bots	IO Archive, ASPI, Transparency Reports

Frameworks to Think Through

https://disarmframework.herokuapp.com