



THE BROOKINGS INSTITUTION

THE CURRENT

“Parsing disinformation in the Israel-Hamas conflict”

Wednesday, October 25, 2023

Host: Fred Dews, Office of Communications, Brookings

Guest: Valerie Wirtschafter, Fellow, Foreign Policy, Artificial Intelligence and Emerging Technology Initiative, Brookings

[music]

DEWS: You're listening to The Current, part of the Brookings Podcast Network. I'm Fred Dews.

The war between Israel and Hamas militants operating out of Gaza has produced horrific images and also misinformation and disinformation about actions on both sides. This myths and disinformation spreads through social media like X, the former Twitter, and confuses our understanding of what's happening. And worse. To help us understand misinformation and disinformation in the current conflict between Israel and Hamas, I'm joined by Valerie Wirtschafter, a fellow in foreign Policy and the Artificial Intelligence and Emerging Technology Initiative at Brookings.

Valerie. Welcome to The Current.

WIRTSCHAFTER: Thanks for having me.

DEWS: So, the broadest sense of this conversation is misinformation and disinformation. Before we get into the specifics of the Israel-Hamas conflict, can you explain if there's a difference between those two terms?

WIRTSCHAFTER: Yeah, I think so. The main difference really is in thinking about intent. Maybe accidentally spreading something would be, I say, traditionally defined

as misinformation, whereas intentionally with some objective, whether it's to sow chaos or maybe makes money, that would be a bit more on the disinformation side.

DEWS: Okay. And I think both of these are probably at play in the Israel-Hamas conflict. So feel free to speak to whichever you want thing most appropriate. But what are some of the top, let's just call, it misleading stories. I mean, the bombing near a hospital in Gaza City comes to mind, that horrible situation.

WIRTSCHAFTER: That is definitely a top one that I think tripped up media, legacy media like The New York Times. Definitely filtered into the online space as well. The other one that comes to mind is the idea that Ukraine is potentially providing weapons to Hamas militants is, I think, another big one. Some others are that the initial attack on a lot of the kibbutzes around the Gaza area, that was a false flag operation. And then there's the uncertainty also around this narrative of the 40 beheaded babies or the sort of visuals, the extreme levels of brutality.

I think all of those are areas of uncertainty, and areas where, you know, whether it's exaggerations because of sort of incomplete information or a deliberate attempt to maybe vilify a certain side in this conflict. All of those, I think, are definitely in play as well.

DEWS: It kind of feels like whether it's deliberate or not deliberate, that a lot of people react to it in a way that kind of confirms their own prior biases. Would you talk about that a little bit?

WIRTSCHAFTER: Yes. You know, I think people are susceptible to this type of information because of really kind of basic underlying cognitive processes. We have our beliefs. We like to have our beliefs confirmed. This is confirmation bias. And we're going to seek out information that confirms those beliefs or discount information that maybe goes against it. It's very motivated reasoning. And so I think that people want to have their priors confirmed. They have certain opinions, very strong emotional reactions, particularly to this this crisis going on right now. And so that tendency to seek out information that confirms what we already think is, I think, very common here.

DEWS: So when we're talking about deliberate disinformation, I mean, how does that work? Are there are there people just taking advantage of the situation to put out bad information because they know it will have a certain effect in the discussion or in the perception of what's going on? And if so, where are these actors?

WIRTSCHAFTER: So I think there's kind of two answers to that question. One is maybe more on the platform side potentially. We've seen a lot of changes, particularly at X, formerly Twitter, that incentivize viral content that incentivize page views for monetization. So I think a lot of people have especially these sort of verified accounts that, you know, recent research has found that they were responsible for spreading, I think, something in the upwards of 70% of the misleading claims, these sort of blue checkmarks they were historically known as.

But now you ... that verification process doesn't happen for notable accounts, maybe journalists, politicians, companies, things like that. It's it's provided to anybody who

pays a certain fee per month. And then you get a bit of a boost in the algorithm as well.

And so all of these little changes have sort of snowballed into this space where it's profitable to spread viral content. And so what spreads in these contexts is videos of explosions. Maybe they're old clips, maybe they're from video games. We've seen some from video games recently that spread like wildfire across acts. And so that's kind of the monetization side of just your lay user.

Then on the flip side, there's certainly actors that have incentives in maybe portraying one side or the other as particularly brutal or particularly irresponsible. And so that may filter into this sort of exaggeration, especially in these kind of uncertain times when the information is incomplete.

And then, you know, what we don't know at the moment is whether there is sort of more deliberate, state driven strategies. There was the narrative about Ukraine and Gaza and Hamas in the arms deals. Was that a deliberate Russian tactic? We don't know. Research will look into that. I'm sure in sort of the way that these narratives have been echoed to maybe deflect interest in other parts of the world, like the war in Ukraine. And so I think all of these different factors are at play.

DEWS: There's another phenomenon that I first noticed on then Twitter, right when Russia invaded Ukraine last year, and I've seen it again play out in the Israel-Hamas conflict, and that's what's called "OSINT," open source intelligence accounts purporting to analyze in great detail key events on the ground, that that explosion near the hospital in Gaza City was one example of OSINT analyses by lots of different people. Are these reliable ways to learn details of complex events, or can that be just as much misinformation and disinformation as anything else?

WIRTSCHAFTER: I think the answer to that is both depending on the source of that information and the care that they take in their effort. So one of these sort of OSINT groups for a while is Bellingcat. You know, they've done a lot of great work around the invasion of Ukraine, exposing some of the atrocities there, and they rely on these things like photographs posted online, satellite imagery that's made available for free, things like that, to be able to either corroborate or push back against the official narratives.

The challenge, of course, is that in these contexts where speed matters, virality gets payment and the term OSINT comes with some sort of level of credibility. People can capitalize on that type of space and use it and spread content that may or may not be true, that may or may not have gone through a clear vetting process, and they may just kind of fall prey to these same challenges of confirming what they want to share.

And so I think that that that's a really big challenge because OSINT researchers do have an important role to play. But, you know, unfortunately, because of a lot of these challenges in the information space, that term has in some ways become a little bit weaponized or diluted in meaning by virtue of the fact that really the type of care that would make this a very valuable tool can be certainly lacking.

DEWS: Valerie, you've done a lot of research over the past few years on how information travels through various digital media, not only social media, but podcasts. There's some reports on our website that you and a team have done on podcasts that I'll link to in the show notes. But given your work in this area, first of all, I'm curious, I mean, how do you actually do this kind of research? Are you just on social media sites all day long?

WIRTSCHAFTER: No, fortunately not on social media sites all day long. Do not listen to hours and hours of podcasts. That report you're alluding to was about 30,000 episodes of podcasts. And I did not listen to 30,000 podcast episodes. So there's a lot of tools of data science, natural language processing to be able to pass data and to be able to parse text data, to be able to transform audio data into text data so that we can quantify some of these trends at scale. Of course, as part of that, there's always kind of a vetting process, a review process, a manual getting in the weeds and seeing, okay, we have a match for something that resembles maybe a podcaster talking about the Bucha massacre as a false flag. Are they talking about it because they agree with it or are they pushing back against it as a conspiracy theory? And so you always have to get in those weeds. That's very important. But to be able to harness some of these tools is really valuable, I think to be able to speed up the process and be able to look at this space at scale.

DEWS: I think that's really fascinating, the methodological approach that you take. But your average citizen who's consuming social media, say in the Israel-Hamas conflict, doesn't have access to those tools. They just see what they see on their phone or on their laptop. Do you have any suggestions on how people can spot online disinformation, even misinformation generally, but also related to this particular conflict?

WIRTSCHAFTER: Yes, definitely. You know, my suggestions are kind of twofold. I'd say, first of all, if something is really shocking, pause, avoid that kind of knee jerk reaction. Chances are there's potentially context that's missing. That's what we saw in the the strike that happened in the parking lot area around the hospital in Gaza. And so maybe, you know, waiting a little bit for the full picture or the uncertainty around the whole picture to reveal itself.

And then, you know, I know that this is maybe cliché at this point or something a little bit trite, but do your own research. I think you can look for additional reports. So say you see something on social media that seems to be particularly controversial or interesting or, you know, you want to you want to share it or send it on a WhatsApp thread or something like that. Check around, see if other sources are maybe reporting a little bit about this as well. Google searching if somebody is I'm a middle East expert or I'm a fill in the blank expert on this topic, search them. Are they in fact publishing on this topic? Are there sources that are considered reputable that are citing this person? Things like that I think would be really useful as well.

And then if you see images, especially in this world of generative AI, especially when generated images can cause, you know, sort of momentary slips in the stock market like we saw when there was a generated image of the Pentagon and a blast exploding near the Pentagon that was generated. It did lead to a stock market dip. But if you see an image, do a reverse Google image search, where's that image

coming from? If it seems to be all on social media, all at the same exact time, if it's not maybe being used by other sources, that would be considered reputable. Maybe pause a moment and think about those kinds of factors of a generated image that would lead you to think something is potentially false. If it's a people, what are their hands look like? What's the background like? What's the sort of gloss on the image?

All those things just just kind of approaching this type of information with a little bit more deliberate care, particularly given the fact that not doing so or kind of running with an idea can really lead to real world violence potentially. And so I think having that sort of pulse check is really, really important.

DEWS: I also think to myself when I see these kinds of stories come across my social media feeds, that maybe it's not my job or it's not in my brief to share this because I'm not an expert in whatever is happening. I might read it, move on, or even do the deeper research that you suggest, but then just have that knowledge for myself. Like, I don't need to share this, right?

WIRTSCHAFTER: Yeah. And there was there was an op-ed in The New York Times, I think, to basically that effect, maybe you just don't have to hit send. And so I think that that that is also really something to think about in this context, too, is this idea of posting your outrage or posting online your your whatever side of this conflict you're on, You know, maybe having these deeper conversations, doing this type of different research can substitute for the kind of cognitive load that comes with that sort of outrage process.

DEWS: And so, Valerie, zooming out now, are there any policy approaches that that you're thinking about to this kind of misinformation and disinformation, whether from government or maybe approaches that businesses like X should take?

WIRTSCHAFTER: I mean, it's an interesting question. And I think that you're, you know, were part of the reason this is such a challenge is that we're in a moment of fluctuation. There is a clear fragmentation, a little bit of the online space. So with a lot of recent changes at X, which used to really serve as sort of a gateway to a lot of on the ground coverage, you know, the Arab Spring was the sort of momentum that built up around that in terms of public awareness. Twitter was a main driver of that.

And so I think that as these changes, it acts have made the platform a bit less reliable as an information source. People don't really have sort of a consensus on where to go. And so I think that that is a challenge that I think will eventually shake out in some respects is sort of where do these on the ground voices come from?

But in the interim, I think there are basic processes. There's information vetting that happens as part of the journalistic process. And so really leaning on these kinds of. Tools of the trade, I think will be really important.

And then from a policy perspective, you know, I wouldn't dare try and influence any policy changes in X. But I think that they've leaned very heavily on this crowdsourced content moderation challenge. And so in the abstract and in practice, even crowdsourced content moderation is fairly decent at its job. I had a report that looked at particularly community notes and how it's performed since it rolled out and

expanded. And it was a mixed bag. The volume of tweets that received notes got better, but it was still way too small. And I think what we're seeing is that in these sort of rapidly evolving crises, it just cannot substitute for some of the more traditional content moderation approaches. And so for platforms thinking about this space, you know, it's got to be, I think, a mix of a variety of different content moderation types.

And so, you know, I think that that there challenges on multiple fronts, particularly acute given a lot of the shifts in the information space that have happened even in the past year. And also, you know, with respect to researchers and having data access, that has also been curtailed. And so I think all of these things, particularly on the data access, would be very valuable for researchers who want to really investigate the roots of some of these narratives. So that's a long way of saying there's a lot of challenges in this space that I think have sort of become magnified a little bit more. It was already a difficult space, and I think in the past year it's become even more challenging for a variety of reasons. And so there's a lot of work to be done for sure.

DEWS: Well, Valerie, I'm I'm glad that you're doing the work to bring light to this situation, both specifically in the new conflict in the Middle East, but also more broadly. We'll continue to follow your research, Brookings dot Edu and elsewhere. Thanks for your time today.

WIRTSCHAFTER: Thanks for having me.