
DISINFORMATION IN ELECTIONS

How And Why It Works, And What To Do

Contents

- 1 Definitions
- 2 Risk Profiles Of Political Campaigns
- 3 Disinformation In Elections: Framework
- 4 Understanding & Predicting Impact
- 5 AI Accelerating Disinformation
- 6 Proactive Defensive Measures
- 7 Response Strategy
- 8 Response Options

1 Definitions

Disinformation

Disinformation is the intentional creation and/or dissemination of false information. It requires intent – the author or bearer of information must know it to be verifiably false but still choose to present it as fact. Motivations for this include but are not limited to influencing public opinion, maligning an opponent, or creating a benefit for themselves.

Disinformation campaigns often combine accurate and false information. This mix forms a more convincing narrative and makes the truth harder to discern. Although disinformation is not a recognized crime, parts of it can be pursued under fraud, libel, or harassment claims.

Misinformation

Misinformation is spreading false or inaccurate information without the intent to deceive. While the information is still wholly or partially false, its spreaders are unaware. The distinction between disinformation and misinformation thus lies in the genuine belief in false information on behalf of the author or distributor.

Active Measures

Active Measures are operations conducted by foreign Intelligence Agencies to advance the aims of those countries. They can include disinformation, espionage, and clandestine operations, for example, to weaken a target country. Discovered Active Measures aimed to sow discord, influence political outcomes, and erode trust in democratic institutions.

2 Risk Profiles Of Political Campaigns

Active Measures are naturally the most concerning intervention in the electoral process but also the rarest one. Intervention by foreign powers is most likely to target those offices that influence foreign policy. The most widely known example is the Russian hacking of the Democratic National Committee in the presidential campaign of Hillary Clinton. More recent efforts by the Chinese government have also targeted Canadian federal lawmakers, but no evidence exists of targeted efforts below the national level. Indeed, it has long been a concern of disinformation scholars that suspecting a foreign origin whenever encountering disinformation can be harmful to democratic institutions.

Campaigns below the national level are statistically vastly more likely to encounter disinformation of domestic origin. These narratives also more rapidly become mis-

information as they are often more believable than those produced by foreign bureaucracies.

3 Disinformation In Elections: Framework

Disinformation usually emerges as a narrative — a false representation or interpretation of facts. Three distinctions are useful to understand how it interacts with elections.

3.1 Target

The target distinguishes narratives by their focus. Disinformation is overwhelmingly negative in sentiment and thus needs some screen on which to project or attach it. These targets rank from the specific to the general.

Candidate Focus often involves personal attacks or false narratives that aim to undermine a candidate's credibility or character.

Issue Focus involves presenting skewed data or false expert testimonials to sway voter attitudes towards or against specific policy proposals, creating a distorted perception of what the majority of the public supports.

Party Focus may aim to tarnish a party's overall image by linking it to extreme ideologies or misrepresented actions. This includes exaggerating the party's stance on controversial issues, fabricating incidents to highlight incompetence or corruption, or associating it with fringe groups to alienate moderate supporters.

System Focus is the subversion of the entire democratic system to sow doubts about electoral integrity and the legitimacy of nations and governments.

3.2 Source

A useful heuristic is that the more specific a target is, the closer the source. For example, a disinformation narrative about an individual candidate will likely come from local actors. Consequently, many different individual narratives about candidates are created independently from one another in a decentralized fashion. Disinformation narratives aimed at a broader target (such as a party or the electoral system) are more likely to emerge in a centralized fashion. Often, these are produced or popularized by one or a few hubs or individuals. The source of disinformation is directly related to how likely people are to believe it. This is true not just for the original source but also for subsequent outlets that pick up on the narrative and

spread it. Both the number of outlets and the trust that individuals place in these outlets can strengthen the disinformation narrative.

3.3 Method

The method of disinformation is a function of what is said and how it is presented. The content (what is said) is usually defined in relation to the target (what false narrative is believable? What will damage the most?). The how is usually related to the audience. Today, disinformation can take all forms – from text to pictures, audio, and video. Thus, when malicious actors put in the extra work in order to produce a more cost-intensive piece of disinformation (such as video or audio), it is usually in response to a specific audience or more ambitious goals.

4 Understanding & Predicting Impact

There will be disinformation in the 2024 elections. This is a certainty. We are already seeing intentional campaigns begin to unfold and currently expect an escalation closer to the election. But while all disinformation is concerning, not all of it is equally impactful.

The impact of disinformation on elections is defined by its impact on voters. Those who are closely engaged with politics tend to overestimate the impact of any one disinformation narrative. This is because they overestimate how many voters are actually exposed to a narrative and may find it believable. To assess and predict the impact of any one narrative, it is helpful to use the target, source, and method framework. **Nevertheless, the first and most important task is to know the voters.**

4.1 Target

Guiding Question: What do your voters care about?

Disinformation is more effective when it targets something people believe in or are more emotionally invested in. What animates voters the most varies by region and up and down the ballot. For a considerable portion of American voters and, therefore, races, the partisan affiliation of those running for office supersedes the importance of individual attributes. Therefore, for partisan races, disinformation targeting the party or the candidate's affiliation with a party is more consequential. The reverse is true if a candidate runs mostly independently of their party. At that point, the targeting of personal character or background becomes more impactful.

Another factor is how localized the race is. For more local races, disinformation is rarer, but if it occurs, it usually is more closely related to local concerns. Disinformation is more likely if there is a source of local uncertainty, such as significant changes to infrastructure, employment, or government.

4.2 Source

Guiding Question: How believable is the source to your voters?

Assessing the source for affiliation can be helpful. Research shows that disinformation that originates in strongly partisan sources is less likely to leave them. This makes sense as the structuring of segregated ideological space on social media (sometimes called echo chambers) makes it challenging for content to migrate from one to the next.

Equally important is the credence that voters are willing to extend to a source. Sources that are deeply integrated into local life and have been for some time are proven to be more believable. But the partisan identity of a source also influences the openness to it. Often, the most critical question is what sources matter to the independent or undecided voters. This consists of knowing the electorate, which can have a wide range of primary news sources - from TikTok for college students to legacy media for many older groups. However, it is also essential to know which parts of the electorate are the likeliest to, in effect, decide an election and which should be reached first (swing voters).

Finally, the number of sources that are repeating a narrative is significant as it will influence both the likelihood of being exposed to a narrative and the results of a search for more information. Nonetheless, these numbers can be misleading. Research indicates that one highly trustworthy source sharing disinformation is more impactful than hundreds of low-credibility sources. Especially the impact of wholly online narratives is often overestimated as the numbers can seem daunting but the actual exposure of local voters can be small nonetheless. When in doubt, the focus should be on legacy and local outlets and those seen as trustworthy to independents.

4.3 Method

Guiding Question: How is the narrative supported and made to seem real?

In general, some forms of content are more believable than others. Watching fake video footage of an event that never took place is more believable than just reading about it. Therefore, when evaluating the impact of any disinformation narrative,

how the information is presented definitely plays into it. The first step is to analyze whether the disinformation narrative claims to have evidentiary material (i.e. is there video, audio, or pictures purportedly supporting it). If so, it is likely that

- The material does not support the logical conclusion that the narrative claims it does (a picture of a dog is not supportive of the claim a dog bit someone)
- The material is taken out of context (a picture of any dog biting someone is not the same as proving a specific dog did at a specific time)
- The material is fake (for example, photoshopped or AI-generated)

Fake material is obviously the most concerning as it potentially has the highest impact, followed by material taken out of context. Their presence increases the potential impact and, therefore, the necessity of response. They can also be more successfully disproven in a succinct and convincing fashion. This impact is still mitigated by both the reach and trustworthiness of the source but is becoming a more significant concern with the increasing role of AI.

5 AI Accelerating Disinformation

Recent research has shown an acceleration in disinformation by Artificial Intelligence (AI). This acceleration consists of three components that may come to bear: speed, scale, diversification, and message tailoring.

Written disinformation can be produced by AI systems that specialize in generating text at great **speed**. The best-known of these Large Language Models (LLMs) is ChatGPT. While companies have tried to build safeguards against misuse, they can and are actively being circumvented. The introduction of AI-generated text has greatly accelerated the speed at which disinformation can be produced, without compromising significantly on quality. As a consequence, new disinformation operations are already operating at vastly expanded scale.

While our ongoing research documents this development, it is difficult to correctly estimate just how large that increase in disinformation activity is. To get a sense of **scale**: currently, research estimates that around half of all content on the internet is now AI-generated. This is a significant increase in the size of the internet as a whole and especially of those parts of it not generated by humans. While we estimate that only a tiny fraction of that is disinformation, it gives an idea of how much AI has enlarged the scale of available text.

The increase in **speed and scale** of disinformation operations through AI is most noticeable on platforms with low entry barriers. Especially Twitter (now called X) and, to a lesser extent, Facebook are currently experiencing a significant increase in

AI-generated content and automated accounts. Often, these automated accounts are meant to create fake attention in order to exploit algorithms that promote content that is seen by many accounts to other accounts. In the past, these bots were more straightforward to distinguish, as they usually posted very little or only text that other accounts were posting as well. However, AI-generated text has allowed these bot-accounts to post text that is much more difficult to distinguish from humans and separate from what other bot-accounts are writing. This makes bot-accounts much more challenging to identify, compounding the problem.

Another forum where this increase in speed and scale is also noticeable is the number of disinformation websites that pretend to be local news outlets across the United States. These AI-generated sites often steal content from legitimate local news sites and have AI change the wording of the stories to obfuscate the origin. These legitimate stories are then mixed with disinformation stories. The goal is to build trust with local readers and successfully persuade them that the disinformation stories are real. However, our research so far shows neither strategy paying significant dividends. Current efforts are hamstrung by the fact that many Americans still receive their daily news from news outlets, even if they may find these stories through links on Facebook or X. ***Our assessment is that disinformation narratives shared through legacy media such as TV will remain more impactful in this election.***

AI has also driven the **diversification** of producible content with fake pictures, videos, and audio files. The ability to fake videos or pictures of candidates has caused great concern, as “seeing is believing”. Studies have shown that people are more likely to believe an event took place when shown a video about it rather than reading a text. This does not necessarily make a video always more impactful. Whether people trust the source remains a significant mitigating factor. Videos shown on legacy media are still seen as more trustworthy than those found on social media.

Furthermore, current studies show that voters are still more likely to doubt AI-generated content than believe it. However, new and more capable models are constantly being released. The most cutting-edge programs currently used at research universities and companies produce very convincing fake pictures, audio, and videos. Nevertheless, the more readily available programs used in disinformation campaigns so far also make more mistakes in the output they produce. If one has reason to believe that an image or video of your candidate is fake, it helps to write down what may be “off” about it. Focusing on small inaccuracies that AI produces, such as the number and shape of fingers, number of teeth, physics of objects and hair, and expressionless eyes, can quickly expose many fake pictures and videos.

However, newer and more capable models are constantly being released. Additionally, the shorter an AI-generated video is, the lower the chance of a glaringly obvious error. Nonetheless, the barriers to producing such convincing footage (computing power, human review of output) will likely remain high. Our level of concern about more ambitious ideas, such as AI-generated influencers on TikTok, remains low in this election cycle.

Lastly, AI could enable disinformation actors to develop more custom and fine-grained messaging that responds to the audiences' individual preferences. This **message tailoring** process has so far not been seen in any campaign but remains a credible threat. With the proliferation of large datasets on many Americans, actors could combine data mining with message tailoring by AI-generated text. Research has shown that messages tailored to the preferences of their viewers are more convincing. If AI supercharges this by allowing disinformation to respond to personal preferences and biases, it could make each individual disinformation narrative more impactful. This has not been seen in any campaign, but our research is continuously scanning large parts of the internet and actively looking for any indications that this may be taking place.

Despite these developments, the amount to which society is more susceptible to AI-fueled disinformation remains open to debate. In general, the more an electorate gets their news from online sources, the more susceptible they are. However, this also means that for many Americans, traditional news outlets remain mostly effective gatekeepers against AI disinformation. ***The impact of any narratives that are repeated on national television or by local outlets is, with high confidence, going to be greater than that of narratives that remain entirely online.***

6 Proactive Defensive Measures

The most important proactive measure is to know the voters. Understanding what issues and topics matter to them enables gauging the significance, spread, and risk of disinformation narratives. It also enables custom response strategies.

Following this, it is vital to know the media ecosystem the constituency consumes and responds to. What is the media consumed by most people? Where are editorial decisions made? What is the potential for abuse? This knowledge also allows for proactive engagement. People are much less likely to believe in disinformation if they know more about the subject matter in question. In going where people get their news, issues can be engaged early and disinformation weakened by stronger voter literacy.

This is also true for voter literacy concerning a candidate. The more voters know, the less inclined they are to believe in disinformation narratives about the candidate. A good start is to publicize and visualize a candidate's voting record (if they have previously held office) on the candidate's websites, if possible, including short explanations for their vote. This can also work for the candidates' positioning on issues, especially if issue descriptions are linked to concrete facts, preferably using citations.

If a candidate is concerned about their likeness being used to create deepfake images or videos, countermeasures can be taken. Readily available online tools, such as Nightshade, change images in a way that is not perceivable to the human eye but complicates processing them with AI. This protective measure changes the smallest unit of color in the images (so-called pixels) but only in ways perceivable to machines. This makes the creation of AI imagery or videos using the candidate's likeness much more challenging.

7 Response Strategy

Being the subject of disinformation or misinformation is disorienting and highly stressful. The first instinct in such situations is to act and go on the offensive. But this can be detrimental. Responses that have errors or elevate a disinformation/misinformation narrative can actually make a situation worse – despite best intentions. Taking the time to assess whether it is even necessary to respond, understand what drives the narrative, and consider the possible response options pays dividends.

Most disinformation and misinformation narratives spread slower than one assumes. Many people just perceive it differently due to logical fallacies – those narratives which spread with great speed are also likely to spread wider. Therefore, fast-spreading narratives are also encountered more often. Still, many narratives never reach these spreads and do not move at such speed. But since one is much less likely to encounter them, there is a bias to overestimating the speed of any one disinformation narrative. Furthermore, few observers are aware of how long a narrative may already have been in circulation before it reached them.

Hence, while monitoring of a disinformation narrative should be ongoing it is crucial to take the time to fully assess it.

7.1 Assessing Reaction Necessity

A reaction is not necessary in all cases, or at least not immediately. It is challenging to deny disinformation without repeating it. However, repeating it can give a narrative new relevance and credence that it may not have had before. If possible, it is preferable to let disinformation fade into irrelevance.

Assess whether the target of the disinformation is relevant to the particular campaign. This follows the target, source, method methodology to predict its impact. Applying these factors to map the impact of a specific narrative onto a limited electorate yields valuable insights. Monitoring the narrative while preparing a response to be publicized in the event of an escalation is recommended.

7.2 Understanding The Narrative

Understanding the narrative is a multistep process that requires taking the most charitable interpretation of a disinformation narrative and assuming no malicious intent. It is crucial to remember that not all false information is necessarily disinformation. Disinformation implies an intent to willfully lie on behalf of the author or distributor. Consequently, no disinformation narrative is successful as disinformation alone. It must become misinformation in order to spread – people must believe it. So, a debunk must address and thus understand:

- Assuming no ill intent: What is the fact that the author could be getting wrong to mean well but arrive at a false conclusion?
 - How hard would it be to verify this fact?
- What is the argumentative chain of the narrative?
 - What are the facts and falsehoods within it?
 - How do they build on top of one another?
 - What is the internal logic?
- Why does an audience find it believable?
 - Does it fill an informational void?
 - Is it built on an old narrative?
 - Is there a lot of uncertainty or emotion?
- What emotions are activated by the narrative?

Taking the time to answer these questions is crucial. Formulating a debunk requires getting to the heart of the narrative. The questions are tools to get there faster. Once someone believes a narrative, getting them to stop believing is a question of

persuasion. That persuasion is borne out of understanding why they believe it. This is true for both indirect and direct responses. The key to persuasion is to address what makes the narrative believable, be it false fact or emotion.

7.3 Preparing With Accuracy

Making a mistake in a response is usually worse than responding a little slower and more accurately. Mistakes in debunking disinformation give opponents an opening to attack the response and discredit both the message and those communicating it. Research has shown that having to correct a mistake in a disinformation/misinformation response or debunk significantly lowers the response's effectiveness. This drop in effectiveness holds true even for audiences that were skeptical of the disinformation/misinformation narrative to begin with. Even those groups that should be the easiest to convince are at least given some pause by a correction in a disinformation response/debunk.

Many situations will afford responders a chance to move more slowly and thoroughly. Especially when a narrative has not yet traveled a medium that matters to key voters. Whenever possible, monitoring a narrative as it develops and preparing a high-quality response is the best way. The goal is to directly address the factual or emotional basis of the narrative, while citing authoritative sources from academia, government, and private sector.

7.4 Calling Something Disinformation

Not all false narratives are disinformation. Often, it is impossible to prove malicious intent behind a false narrative. In those cases, calling it disinformation would be misleading. But this is not just about scientific accuracy. It can also help with debunking. It is harder for individuals to be convinced of something when they feel they are blamed for it. Calling a narrative disinformation can do just that to people who fell for it. Embracing them and arguing from a place of understanding is much more effective. Giving a narrative the benefit of the doubt can make the response to it more effective.

Whether it is advisable to label something disinformation also relates to what it targets. While issues are removed and more prone to faulty interpretation of facts and thus misinformation, this is less true for narratives targeting a candidate. In response to these, research indicates that candidates who defend themselves aggressively are perceived more positively. Should a candidate's character be the subject of disinformation, a more robust response can be useful.

8 Response Options

If it has been determined that a misinformation/disinformation narrative is relevant to the electoral campaign because of target, source, and method, it is necessary to engage the narrative.

8.1 Indirect Response

This works best if the narrative is so far still fluid and more issue-focused rather than a specific personal attack on a candidate. This is the preferred strategy as it enables counteracting the narrative without repeating it. Repeating a narrative can give it credence and attention. Whenever possible, this should be avoided. Instead, the underlying falsehoods or emotions of the narrative can be addressed without repeating it. This is achieved with an indirect response.

As part of the indirect response, the candidate also starts speaking about the underlying falsehoods or emotions of the narrative on the record – without repeating the narrative. The indirect response allows a candidate to air accurate information and make their viewpoint known. It confronts those who believe in the narrative with a different viewpoint and equips those who may be vulnerable to it with facts. Good opportunities to respond indirectly to a narrative can be created by scheduling an event related to the falsehoods or emotions underlying the misinformation/disinformation narrative. Such events give the candidate both the visibility and opportunity to offer factual information and increase the chance of repetition in media.

The key is to prepare messaging that engages the disinformation narrative's central claims without repeating it. This route is preferable, as it lends no credibility and publicity to the disinformation narratives. It, however, has the risk of potentially being less effective than a direct response and likely being a greater allocation of resources and planning to set up.

8.2 Direct Response

The response should begin by laying out the facts. Without referencing the disinformation/misinformation narrative, it should summarize the facts of what is actually true. The goal is to fill the gap that disbelieving the narrative would lead to. A helpful way to think about this is a one-sentence statement about reality. Reference only verifiable information; if something is an interpretation, state this.

Then, the response should proceed with working through the disinformation/misinformation narrative and debunking it. The goal is not to stick with simple denials or refutations. Instead, it is about following the questions outlined in "Understanding the narrative". This is a roadmap. The response will be effective if it isolates the falsehoods underlying the narrative and disproves them. The response is convincing by understanding the emotion that underlies the disinformation/misinformation narrative and acknowledging it. Showing understanding for those who believe it without malicious intent is critical. Success in this is more persuasion than education. An excellent way to think about it is as debating the narrative. While doing so the narrative should be referenced only once. In follow-up questions, repeating the narrative is to be avoided.

Finally, the direct response closes by rephrasing the opening statement. Doing so offers the logical contrast between the disinformation/misinformation narrative and response. Closing statements also have the highest retention rates. Thus, it is important not to finish with the disinformation/misinformation narrative.

And then repeat, repeat, repeat. To decrease the effects of disinformation and misinformation narratives, repetition is key.