FABLE Framework of Misinformation Harms

A Structured Response to Misinformation as Harm

In fact checking, a critical need is prioritization. How can fact checkers be supported in their decision making to decide which pieces of information, when false, may have more negative impacts in comparison to others? Five major dimensions can help determine the potential urgency of a specific message or post: **(Social) Fragmentation**, **Actionability**, **Believability**, **Likelihood of Spread**, and **Exploitativeness**.

Use the suggested Key Questions (highlighted in blue) if a quick assessment is needed. More about how this framework was constructed can be found here: https://www.artt.cs.washington.edu/analysis-framework-online-misinformation-harm and at https://doi.org/10.1145/3641010.

FABLE Framework

Dimension 1: Social Fragmentation

A piece of misinformation could have indirect, societal, and accumulative effects.

Therefore, a piece of misinformation is more harmful the more that it addresses or is part of societal and community relationships over time.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire

There are no short cut questions for this category.

Social Fragmentation Questions	Y	N	?
Does the message fit into a larger story or argument, for example about how the world works or how people think? *A larger narrative means that there is story touching the information that crosses platforms and time. This may include stories about communities, race, political parties.*			
Does the message question trust in or the functioning of public institutions?			
Does the message question trust in or the functioning of the scientific community as a whole?			
Does the message question the functioning of or trust in news sources/the media in general?			
Does the message question the trustworthiness of other people in general within a community or society?			
In a democratic country where there are elections, does the message directly attack the election process?			
Subtotal*			

Fragmentation	Actionability	Believability	Likelihood of Spread	Exploitativeness	TOTAL
Dimension 1 Subtotal	*Dimension 2 Subtotal*	*Dimension 3 Subtotal*	*Dimension 4 Subtotal*	*Dimension 5 Subtotal*	

**"Don't Know/Not Applicable" (?) can be subtracted from Total "Yes" (Y) answers to gain a relative magnitude of urgency across different pieces of content.*

FABLE Framework

Dimension 2: Actionability

A piece of content is more harmful the more that it spurs actions that directly cause harm.

Therefore, a piece of misinformation is more harmful the more that it spurs direct action.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire.

Use the Key Questions if a quick assessment is needed (highlighted in dark blue).

Actionability Questions	Y	N	?
Does the message content include an explicit call to action?			
Does the piece of content incorporate coordination efforts, such as dates/times or other arrangements for follow-up?			
Does the message provide a name or otherwise any identifying information about an individual, an address, or a place of work in such a way that people might be directly harmed?			
Does the message content include a tone of urgency or mention of time sensitivity?			
Does the message content include any threats of violence?			
Does the message lay blame or cast aspersions or hatred on a particular group, such as a particular religion, gender, sexual orientation, race, country, or culture, that has been harmed in the past by the audience of the content?			
Does the message invoke a sense of injustice or moral outrage, including on behalf of a vulnerable individual or group such as children or women?			
Does the direct target or current audience members directly addressed in the message have a recent history of taking actions that cause harm?			
Is this message associated with/similar to other messages that are also actionable?			
Subtotal*			

**"Don't Know/Not Applicable" (?) can be subtracted from Total "Yes" (Y) answers to gain a relative magnitude of urgency across different pieces of content.*

FABLE Framework

Dimension 3: Believability

A piece of misinformation is more harmful the more believable its message is to a specific community.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire.

Use the Key Questions if a quick assessment is needed (highlighted in dark blue).

Believability Questions	Y	N	?
Is there a lack of high quality information that is publicly accessible and is refuting the message's claim?			
Is there a lack of consensus on the part of experts regarding the claim?			
Does the message fail to include external citations, links, or language about evidence to support its claim?			
Does the message contain richer formats as part of its evidence that lay people consider to have low falsifiability?			
Is the message written or communicated in a personal or persuasive tone?			
Does the message make reference to the broad believability of the claim or topic?			
Does the message appeal to a specific community identity by mentioning a shared set of values or beliefs?			
Does the poster and/or organization/outlet have a noteworthy number of social media/community followers?			
Is the content published by an organization/outlet with uncertain editorial control (e.g., is not a recognized news publisher)?			
Does the poster have credentials that represents some kind of expertise?			
Is the content posted by an imposter individual or counterfeit outlet that could successfully pass as a different person/account based only upon a quick glance?			
Does the content have the graphics and styling of a legitimate news agency or mainstream information source?			
Subtotal*			

**"Don't Know/Not Applicable" (?) can be subtracted from Total "Yes" (Y) answers to gain a relative magnitude of urgency across different pieces of content.*

FABLE Framework

Dimension 4a: Likelihood of Spread

A piece of harmful content is more harmful the more places it appears on and the more people who are exposed to it.

Therefore, a piece of misinformation is more harmful the more places and people are exposed to it.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire.

Use the Key Questions if a quick assessment is needed (highlighted in dark blue).

	Likelihood of Spread Questions	Y	N	?
WHO is spreading?	Do the people or entities who are spreading the piece of content have a broad reach (size of following on social media, "influencer," presence on TV or other news media)?			
	Are the people or entities known to be repeat spreaders of questionable information?			
	Is there evidence of coordination activity (whether bot/automated or not) to encourage spread?			
WHERE is it spreading?	Is the content publicly accessible (posted on a public platform, addressable URL)? Is the content posted on a popular platform?			
	Is the content spreading on multiple platforms?			
	Does one of the platforms upon which the content is shared have tools to support amplification (e.g., reshares, algorithmic feeds, recommendation engines)?			

...Urgency Dimension questions continue on next page

FABLE Framework

Dimension 4b: Likelihood of Spread

A piece of harmful content is more harmful the more places it appears on and the more people who are exposed to it.

Therefore, a piece of misinformation is more harmful the more places and people are exposed to it.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire.

Use the Key Questions if a quick assessment is needed (highlighted in dark blue).

**"Don't Know/Not Applicable" (?) can be subtracted from Total "Yes" (Y) answers to gain a relative magnitude of urgency across different pieces of content.*

	Likelihood of Spread Questions	Y	N	?
CHARACTERISTICS of the message	Does the message make direct appeals to audience members that it is in their financial, political, or social interest to spread the content further?			
	Does the message directly call audience members to share the content further?			
	Is the tone of the content striking enough in ways that encourage sharing?			
	Does the content contain an image, audio-clip, or other richer formats that are easy to remember, visually or aurally arresting, or seems interesting to share?			
	Does the message impart a sense of exclusivity or novelty ("breaking news")?			
	Are there hashtags associated with the message?			
	Is the message difficult to fact-check or prove false?			
	Is the message related to a current event or a topic that is being reported on actively by many news outlets?			
	Subtotal*			

Don't forget the questions from page 4!

FABLE Framework

Dimension 5: Exploitativeness

A piece of misinformation is more harmful the more the message seeks to exploit a group's weaknesses, including a lack of resources.

Mark a "1" for each answer and tally them up.

Suggestions for hints on how to answer these questions can be found at https://www.artt.cs.washington.edu/online-misinfo-harms-questionnaire.

Use the Key Questions if a quick assessment is needed (highlighted in dark blue).

Exploitativeness Questions	Y	N	?
Does the message directly address or reference children or use language aimed at a younger audience?			
Does the message introduce a degree of fear or feelings of uneasiness?			
Is the message content complicated to understand?			
Does the message directly address or reference elderly community members, or discuss topics aimed at them?			
Does the message directly address or reference military veterans, or discuss topics aimed at them?			
Does the message make mention of a reader's feelings of isolation?			
Does the message make mention of a reader's feelings of powerlessness?			
Does the message make mention of a reader's feelings of disenfranchisement?			
Is this message being shared by within an online group, community, or thread that has a recent history of discussing conspiracy theories or viral misinformation?			
Is the language of the intended audience neither a UN language (English, French, Spanish, Mandarin Chinese, Russian) nor on the top 5 list of most popular languages?			
Is the message presented in a region where the local context might amplify its harm?			
Subtotal*			

** "Don't Know/Not Applicable" (?) can be subtracted from Total "Yes" (Y) answers to gain a relative magnitude of urgency across different pieces of content.*