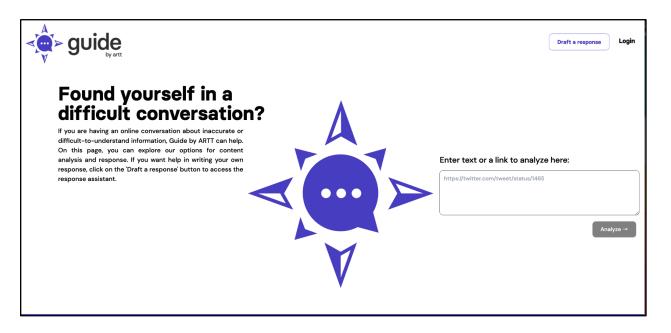
The ARTT Guide User Manual

Welcome to the ARTT Guide User Manual. It is designed to introduce first-time users to the tool's various functionalities, and to make its development transparent.



Section 1: ARTT Tool Quick Start

ARTT Guide is designed to help users assess and respond to posts and articles that include contentious or controversial topics. The Quick Start sections will explain how to access the ARTT Guide's Analysis and Response Features.

Getting Started	2
How to Analyze	3
How to Browse Response Options	5
How to Draft a Response	7

Section 2: More Information

This section describes the Analysis, Response and Community Resource Directory in greater detail.

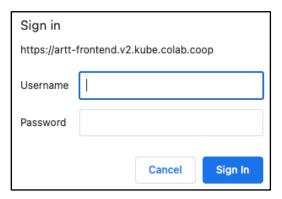
Analyses in Detail	. 1C
Response in Detail	. 15
Community Resources in Detail	. 16



Open the ARTT Guide on a desktop browser: Google Chrome, Firefox, Edge, or Safari. We do not recommend using a phone or tablet.

Logging In: There are <u>two steps</u> required to access the tool for the first time:

1. When you click the tool link, a pop-up window will appear. Please enter the username and password you received in your introductory documents.



2. Once you have accessed the tool, you will then need to log in to your unique tester account.

Please log in at the **top right** corner of the tool, using your assigned User ID and password.



You'll know that you're logged in when your User ID has replaced the 'Login' button in the top right.

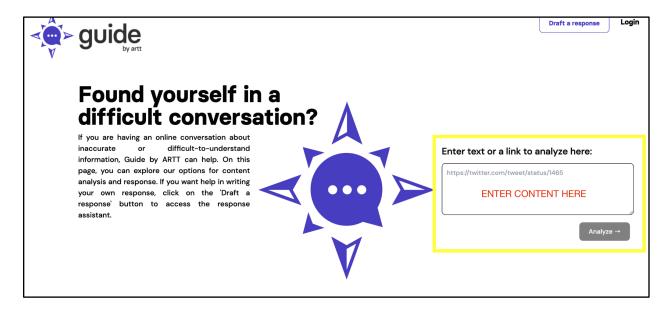


Once you are logged in, you will stay logged into unless you intentionally log out.

You are now able to use the ARTT Guide! You can always return to this homepage by clicking the ARTT logo in the top left.

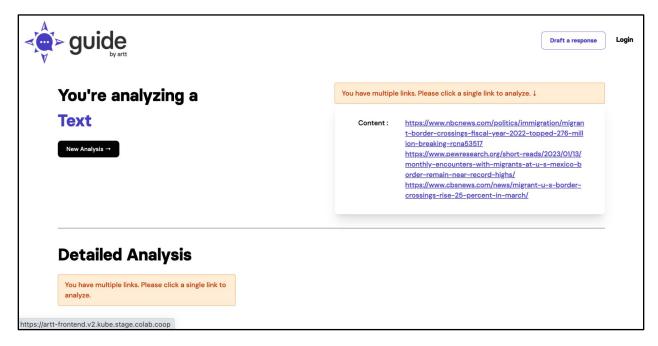


You can analyze content directly on the ARTT home page. To do so, paste text, a link to an online article, a Twitter post, or a Reddit comment into the box.



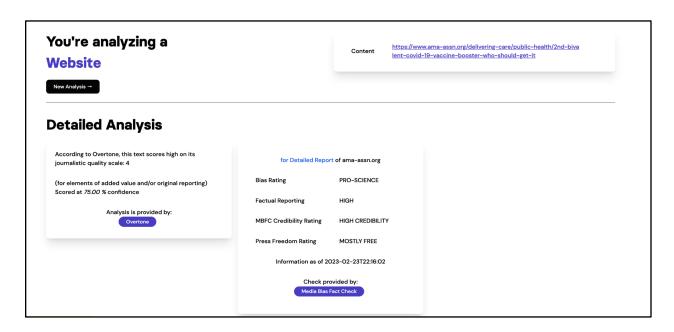
Depending on what you paste in the box, the tool will take the following actions:

- If you enter only text, then that text is what will be analyzed.
- If you enter a single URL, then the content from the URL will be analyzed.
- If you enter text and a URL, then only the text will be analyzed. You will be prompted to click on the link to run a second analysis on the URL. This analysis will open in a new window.
- If you enter multiple URLs, then you will be prompted to select a single link.





After you click "Analyze," you will be shown general information on your entry. The '**Detailed Analysis**' section shows cards with results the different tests run on the content.



Depending on the content, these cards may include:

Analysis Type	Source of Analysis	Basic Description
Source Quality Check	Media Bias/Fact Check	Offers bias, credibility, press and freedom ratings of source according to a methodology defined by Media Bias/Fact Check.
Sentiment Analysis	VADER	Rates polarity and intensity of a post. Scores range from -1 (very negative) to +1 (very positive).
Journalistic Quality	<u>Overtone</u>	Rates journalistic quality of the content (not the source). Scores range from 1 (less depth/low informational value-add) to 5 (high depth/high informational value-add). The confidence percentage represents Overtone's confidence in the score's accuracy.
Twitter Bot Detection	Botometer	Gauges whether a twitter account is likely to be a bot or not. Scale is from 0 (most human-like) to 5 (most bot-like).
Wikipedia	ARTT, in	For Wikipedia articles only. This label provides



Article Assessment	collaboration with Wikimedia DC	information to help readers assess the quality of the article.
Misinformation Harm Potential	ARTT	This is a prototype feature. For now, it will highlight if the content has an explicit call to action (actionability) or if the content targets the vulnerable group of children (exploitativeness).
Fact Checking	Google Fact Check	If a specific claim has an existing fact check, users will see a list of matching claims and the corresponding fact checks.

You can learn more about these analysis metrics in the "More Information" section of this document, <u>found here</u>.

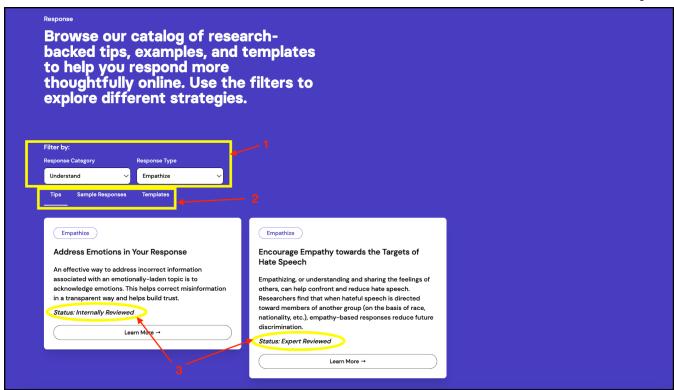
Below the Detailed Analysis, you will find ARTT's "Community Resources on Vaccines." This resource is a directory of ARTT-reviewed vaccine and vaccine-related information. You can filter sources by target audience. On the right, you will find other resources that may be helpful. To learn more about this <u>resource</u> <u>directory</u>, <u>click here</u>.

Community Resources on Vaccines Search	Cleveland Clinic Source: https://my.clevelandclinic.org/landing/covid-19-vaccine Status: Accepted	Other Resources Wikipedia Perennial Sources list WHO Vaccine List Wikipedia Reliable Sources list on Vaccines
Filter sources by community		

How to Browse the ARTT Response Catalog

If you scroll down on either the home page or the detailed analysis page, you will see ARTT's response **Tips**, **Sample Responses**, and **Templates**. These tools aim to help you craft a response and learn about the elements of successful responses to comments, posts, or online statements.





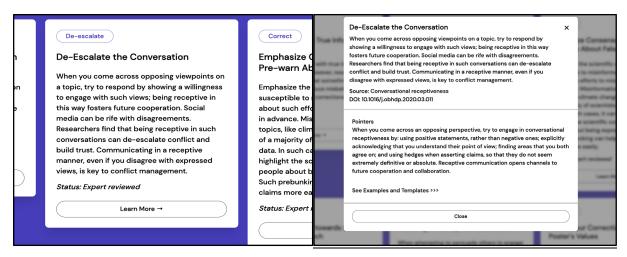
- 1. You can filter response options by "Response Category" or "Response Type." Learn more about these response options in the "More Information" section.
- 2. Toggle to show all of ARTT's Tips, Templates, or Sample Responses.
- The 'Status' indicator at the bottom of each card tells you that every Tip, Template, and Sample has been reviewed to ensure quality:

<u>Under development</u>: Undergoing review by the ARTT team <u>Internally reviewed</u>: Reviewed by the ARTT team <u>External</u>: Out for review or revision by an external reviewer. <u>Expert reviewed</u>: Undergone a second, external expert review *Note: All ARTT tips will eventually undergo external review.*

Tips tell you specifically how you can respond to a piece of content. Click 'Learn More' on any tip card to see its full explanation as well as ARTT-designed pointers.

Note: Tips aren't equally effective in every scenario. Refer to a tip's sources for more information about its testing contexts.





Sample Responses are complete examples of the response tip. If you open a card and select the 'See Examples and Templates' link, you will be directed to a page specific to the relevant tip. You can then scroll down to 'Explore Sample Responses' and 'Explore Templates.'

Templates you can use to copy and paste as a starting point for your responses. You can learn more about each template by clicking on the links under 'See associated tips.'

You can learn more about these response options in the "More Information" section of this document; click here.

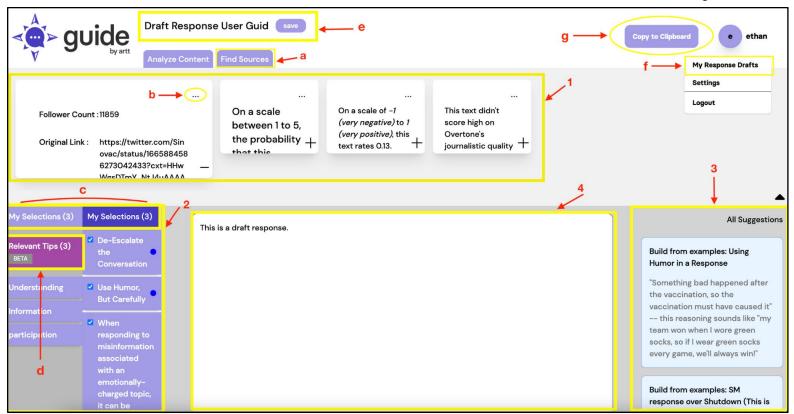
<u>Drafting a Response</u>

To draft a response, click the 'Draft a Response' button at the top of the page. This will bring you to a page where ARTT's analysis and response catalog are harnessed in a tool where you can write a unique response or comment.

Note: If you aren't yet logged in, you will be directed back to the login page first.

The 'Draft a Response' page has four sections: an 1.) **Analysis & Sources Toolbar** at the top, a 2.) **Response Selection Sidebar** on the left, a 3.) **Suggestions Sidebar** on the right, and a 4.) **Drafting Area** in the middle.





- 1.) Analysis & Sources Toolbar You may wish to respond to a specific message or piece of information. Click the 'Analyze Content' tab to paste text or a URL you wish to perform an ARTT analysis on, covered in the previous part of this manual. The 'Detailed Analysis' information will be broken down into cards that are stored on the toolbar.
 - a) If you want to search for a specific source, click on the 'Find Sources' tab to search the sources from ARTT's <u>Community Resources on</u> Vaccines.
 - **b)** You can delete the cards in your toolbar by clicking on the three dots in the top corner of each card.
- **2.) Response Selection Sidebar** Here you can explore the different tips, samples and templates available through ARTT, organized by the goal of your conversation.

You can select a Tip from this sidebar, and if there are any associated samples and/or templates, they'll show up in the Suggestions sidebar.

c) My Selections: Tips that you have selected will be shown in the 'My selections' tab.



- d) Relevant Tips [BETA feature]: Tips that are thought to be relevant based on the analysis and source results will show up in the "Relevant tips" tab. Blue (strong) and yellow (tentative) dots also appear to the right of the tips to indicate the level of possible relevance. This feature is experimental.
- **3.) Suggestions Sidebar -** Based on your selections from the Response-type sidebar, you will see on the right a list of ARTT's relevant sample responses and templates. You can click to paste these into your drafting area.
- 4.) Drafting Area This space is for you to draft a potential response.
 - e) You can title and save response drafts above the Analysis Toolbar.
 - f) View your saved drafts by clicking on your User ID
 - g) Ready to respond? Click 'Copy to Clipboard,' and paste your response into any external website or application.



More Information

This section describes the Analysis, Response and Community Resource Directory in greater detail.

Analyses in Detail

The Analysis feature of the ARTT Guide aims to provide the user with information about content shared online. ARTT utilizes a variety of tools that highlight different dimensions of content, such as accuracy or emotion. Some of these tools are provided by external third parties, and others have been created by ARTT and ARTT partners.

Each analytical tool has a unique rating system and methodology. While basic information about each tool is included in this section, users are encouraged to learn more about the tools on their own using the links provided.

Think of the results as one might find from a Google search: the results from the analysis are not meant to definitively tell the user if the content is good or bad. They are intended to provide the user with different lenses, from different providers, through which to view the content. With these insights, users can be additionally equipped to make their own judgment about a piece of content's quality and reliability.

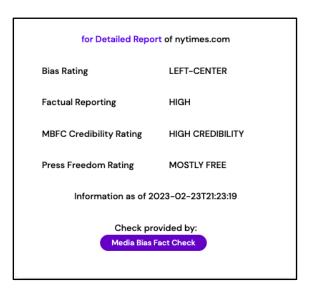
The analysis features of the ARTT Guide include the following tools:

- Source Quality Check
- Sentiment Analysis
- Journalistic Quality
- Wikipedia Article Assessment
- Misinformation Harm Dimensions
- Google Fact Check

Source Quality Check

Provided by: Media Bias/Fact Check (MBFC)

What is it: If your content is an article from a news or media outlet, this tool will give information about the article's source. It will give you different ratings that MBFC has generated or collected, such as: level of general bias, reputation for factual reporting, the MBFC Credibility Rating, and the Press Freedom Rating for the source. You can read more about their assessment methodology here.





Why is it helpful: One way to determine the likelihood that a piece of content is credible is to evaluate its source.

Sentiment Analysis

Provided by: VADER

What is it: This tool measures the polarity and intensity of a piece of content, and gives it a score. It ranges from –1 (very negative) to 1 (very positive). It works optimally on short pieces of text, such as social media messages.

On a scale of *-1 (very negative)* to *1 (very positive)*, this text rates 0.40.

Analysis is provided by:

VADER

Why is it helpful: By gauging the type of emotions expressed in a piece of content, users can see how that content is being presented. Sentiment can be an insightful indicator, because studies have shown that language meant to deliberately evoke an emotional reaction (in particular, an extreme negative emotion, like anger or fear) can make certain messages more persuasive and shareable.

Note: Currently, ARTT Guide treats scores ranging from -0.05 to 0.05 as neutral (based on this paper).

Journalistic Quality

Provided by: Overtone

What is it: This tool uses an algorithm to rate the editorial text, according to the presence or lack of journalistic signals: original reporting, good sourcing and citations, meaningful analysis, and exploration of ideas. The score ranges from 1 (indicating less depth or low informational value-add) to 5 (indicating high depth or high informational value-add).

If a piece of content scores a 3 or less, the L card will simply state: "This article did not score high on Overtone's journalistic quality."

According to Overtone, this text scores high on its journalistic quality scale: 4

(for elements of added value and/or original reporting)
Scored at 78.00 % confidence

Analysis is provided by:

Overtone

This article didn't score high on Overtone's journalistic quality scale

Analysis is provided by:

Overtone

Overtone also includes a 'Confidence Score' based on how accurate they believe their analysis to be. The range is from 0 (not confident) to 100 (very confident).



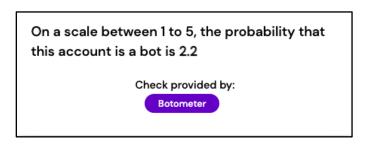
Lower confidence scores indicate either Overtone's unfamiliarity with the type of text provided, or the possibility that the text may have received a different score.

Why is it helpful: This tool assesses the content via a journalistic lens. The score does not take the source into account; it only assesses the quality of the editorial content. While the tool will score all types of content, because the score depends on the presence or lack of journalistic signals, this is best used in the context of a journalistic article. (For example: reference pages on sites such as the CDC may not receive a high journalistic score according to Overtone analysis.)

Twitter Bot Detection

Provided by: Botometer

What is it: This tool assesses the likelihood that a Twitter account is a bot. The score ranges from 0 (most human-like) to 5 (most bot-like).



Why is it helpful: This can help the user decide if a statement or tweet is worth responding to in the same manner as if the message is from a human.

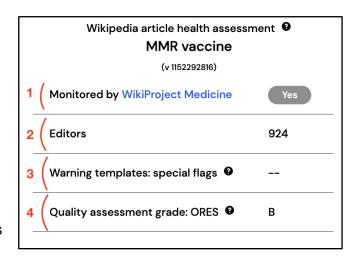
Wikipedia Article Health Assessment

Provided by: ARTT in collaboration with Wikimedia DC

Note: On this card click the symbols to learn more about each section.

What is it: This tool provides a basic assessment of a Wikipedia article. It will check for the following items:

 Wikiproject Medicine Monitoring (Yes/No): Wikiproject Medicine is a group of medical professionals who are dedicated to ensuring article accuracy



- 2. Number of Editors: Tracks how many edits a page receives and how many editors have contributed.
- **3.** Warning Types: Some articles are flagged with a warning. These flags are designed by the publicly-sourced <u>Wikipedia Cleanup</u> project, designed to catch mistakes in articles and to make sure that important and influential



pages stay up to date. ARTT searches for 8 important information reliability warnings:

- Articles lacking reliable references
- Articles needing context
- No Original Research
- Neutral Point of View Disputes
- Articles needing clarification
- Articles that may contain original research
- Articles with undisclosed paid content
- Articles needing rewrite
- **4.** ORES Wikipedia Content Assessment: ORES is a scoring system created and used by the Wikipedia community to assess the quality of an article. ARTT will indicate if the article has one of the following key scores:
 - FA (Featured Article)
 - FL (Featured List)
 - A (Very useful to readers)
 - GA (Good Article)
 - B (Useful for the Casual Reader)

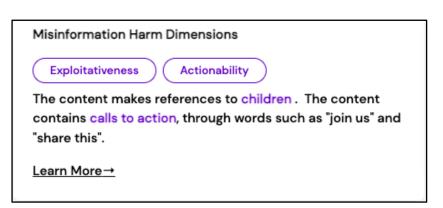
Why is it helpful: Users should look at this assessment as one would a "nutrition label. It is a way for users to quickly gauge the overall health of a given Wikipedia article.

Misinformation Harm

Dimensions

Provided by: ARTT with the <u>UW</u> Social Futures Lab

What is it: This tool gives insight into how harmful content has the potential to be. This is a prototype feature, and still in development. Currently, this card will trigger if the content has an explicit call to



action (actionability) or if the content targets the vulnerable group of children (exploitativeness).

Later versions of the ARTT Guide will aim to implement indicators within the following <u>five dimensions</u>:

1. Actionability: The indicators in this section measure whether the misinformation is actionable in a way that could elicit directly harmful actions, particularly physical ones.



- 2. *Exploitativeness*: This dimension recognizes that factors can contribute to the target audience's vulnerability to misinformation, ranging from emotional manipulation to a lack of available resources.
- 3. *Likelihood of spread*: A piece of misinformation is more harmful the more platforms and people are exposed to it.
- 4. *Believability*: This dimension's measures are related to topics in which either authoritative consensus is difficult to achieve or is affected by the perceptions from a specific community.
- 5. Social fragmentation: This dimension measures how deeply embedded the content's narrative is to the community's existing context and history.

Why is it helpful: These actionability and exploitativeness triggers could help a user identify when pieces of content, if false, may have more negative impacts in comparison to others.

Regarding actionability: A piece of content that is harmful becomes more harmful when it spurs direct action. Therefore, a piece of misinformation is more harmful the more that it spurs direct action.

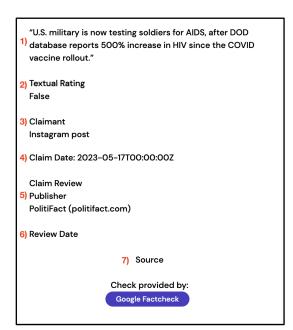
Regarding exploitativeness: A piece of misinformation is more harmful the more the message seeks to exploit human or a group's weaknesses, including a lack of resources.

Fact Checking

Provided by: Google Fact Check

What is it: This tool will show fact checks found on a specific topic. These results are pulled directly from Google's <u>Fact Check Explorer</u>, and will contain websites specializing in fact-checking and other publishers that meet Google's guidelines. Multiple fact check cards may show up from a single analysis, especially if it is a pervasive or controversial claim. In the ARTT Guide, each card will show the following:

- 1. Top statement (Claim): The topic in question
- 2. Textual Rating: The accuracy rating or explanation from the fact-check source.
- 3. Claimant: Whoever made the claim
- 4. Claim date: The date when the claimant made the claim
- 5. Claim Review: The type and name of the source that provided the fact-check
- 6. Review date: The date when the fact-check was published





7. Source: A direct link to the claim review source. Click "Source" to read the full article and learn more.

Note: This card will only show up if there is an existing fact check.

Why is it helpful: This is a quick way for users to see what existing fact checks have already been done regarding a specific topic or claim. It can potentially help users respond to certain claims with information that is reputable or well-sourced.

Response Types in Detail

These response types are context dependent and not exhaustive. Users should judge what response type is best to use in each specific situation. ARTT's research on response methods is ongoing, and is updated regularly in its <u>public ARTT</u> Response Catalog.

Understand is a response type that seeks to comprehend and consider the other person. This is an acceptance of peoples' emotions and perspectives, even when they are contrary to fact.

- Listen By listening silently, participants can understand more about whether
 to respond or how to respond. It is a critical part of a trust-building exchange,
 especially in situations where one is thinking about the possibilities for longer
 term dialogue or engagement outside the immediate message being
 discussed.
- 2. Empathize An identification with someone else on an emotional level. Empathizing is a key mode of responding in conflictual exchanges where resolution or a transformation of the relationship is the goal.
- 3. *Take Perspective* By doing this, people can identify another person's intentions and needs even though they may not agree with them, which may reduce impasses and decrease discrimination.

Inform is a response type that focuses on the quality of information by scrutinizing it, and by pointing out the importance of due diligence to both the author you are responding to and to any other readers.

- Correct One goal may be to correct the speaker about a specific issue, such as climate change or vaccination. Another goal might be to equip the speaker with general skills to identify inaccurate information.
- 2. *Co-verify* This is a method that has been observed in practical interventions in which someone offers to undergo source evaluation and fact-checking processes in tandem with someone else.
- 3. Encourage Healthy Inquiry To encourage healthy inquiry is to help others ask questions of the information they are reading, such as "What do other sources say?" or "What's the evidence?" Being able to critically evaluate information by not immediately believing new claims is an important skill.



Connect is a response type where you actively join the conversation in a tactful way, and seek to strengthen connections to others.

- De-escalate A reduction of hostilities between different individuals or groups. This is an overarching goal of efforts in conflict resolution or transformation.
- 2. Share This response implies a deeper engagement in a conversation. Sharing one's own story is one way that people explain their reasoning through their own personal experience of navigating a difficult decision.
- 3. *Invite Sociability* Sometimes it helps to remind people of the ways that we're connected to one another, whether a shared desire for accuracy or a commitment to moral values. These tips focus on the interpersonal bonds that exist among us.

No Response - Sometimes a valid response type is to not engage at all, such as if you realize that a discussion has stopped being productive, or no longer offers any opportunity for progression or change, or is not one that you are equipped to have.

Community Resources in Detail

What is the "Community Resources on Vaccines" directory?

The "Community Resources on Vaccines" directory contains a selection of question-and-answer sources for vaccine-related information that have been vetted using a set of standard criteria. These resources have captured questions and concerns that are not always readily available or addressed by larger organizations, such as the WHO, enabling them to better connect with different groups and audiences.

What kind of sources are included in this directory?

The directory contains high-quality, question- and- answer resources developed by many community leaders that are in alignment with authoritative guidance on vaccines. Each source goes through a vetting process that asks questions about its credibility, competence, and relevance - questions that are based both on current research in trust and trustworthiness, and existing source vetting standards from organizations like the Vaccine Safety Net (VSN).

Are the sources included here exhaustive?

The sources included here provide a sample of sources that address various vaccine-related questions and concerns from different communities. This directory should be considered as more of a snapshot of resources and is not one intended to be a comprehensive list of sources for a given concern or community.

How is the ARTT tool using this directory?



The primary purpose of this directory is to integrate the information from the sources into the ARTT Guide software, as part of the "Analysis Results" area of the tool. You might see a suggested trusted resource as part of your analysis results for a given source.

How were the resources in the "Trusted Resources Directory" assessed? What criteria were used?

Each source here included in the directory was assessed according to the criteria in the following three dimensions: integrity, competence, and affinity. All sources included in this directory met a majority (at least 50%) of the criteria for competence, and 50% of the criteria in at least one other dimension. See the table on the next page for a comprehensive list of all the vetting criteria.

Criteria Area	Focus of Criteria	Key Question	Additional Guiding Questions
Criteria 1: Integrity of Source	Assess a source's commitment to transparency, integrity, and adherence to professional standards.	Does the source take steps to demonstrate why its information is worthy of being trusted or believed?	 Is the mission and purpose of the website clearly stated? Is there clear disclosure of the ownership/source of the website? Is there transparency in the sponsorship or funding of the website? Is there a way to contact the sponsor or author of the website? Is the website professional in appearance?
Criteria 2: Competence of Source	Assess the ability of the source to convey accurate information, with emphasis placed on whether the source is a member of the Vaccine Safety Net	Does the source demonstrate competence and consistency in scientific expertise around vaccine information?	 Is there a citation or a clear statement of the sources used for scientific, medical and health information? Does the majority of the information provided align with the latest from the WHO and members of the Vaccine Safety Net? Does the information have a clearly marked



			 publication/last updated date? Does the source promote good immunization practices? If including multiple sides of a controversial issue, does it do so in a fair way? Is there internal consistency in the answers these sources provide?
Criteria 3: Affinity of Source	Assess any relevance the source has toward a particular group.	Does the source build affinity with specific communities by taking concerns seriously and establishing connections with them?	 Is this source recommended or run by figures with ties to relevant communities/trusted local community sources? Does the source address questions or concerns that are specific to a given community or demographic? Is the source a non-profit/community-based-organization (CBO)? Do all the source's answers thoroughly address questions and concerns of the community?