Description and Learnings

This dataset was produced through a collaboration with the NSF-funded ARTT project (led by Hacks/Hackers) and Overtone. The goal was to create a dataset of vaccine-related articles, pulled from a wide variety of news media sources, with associated scores based on their journalistic quality. The scores were provided through Overtone's algorithm, and range from one (low-quality or low informational value add) to five (high-quality or high informational value add). Unlike many of the datasets in this field, the quality scoring is solely based on the content of the article itself, not on the reputation or publication history of the outlet or author. Therefore, this collection of articles provides a useful insight into the range of quality vaccine reporting over the course of one year.

Exploration

The creation of this dataset provided an opportunity to see how a news-oriented, AI-based evaluation service focuses on the topic of vaccines. As explained in the Dataset Curation Methodology section, the research team used a variety of sources to collect articles. They included traditional journalism sources (news and news-leaning websites), as well as non-journalistic sources of vaccine information, such as governmental websites, healthcare and NGO websites, and medical journals. Given the algorithm's focus on editorial content, as opposed to other metrics such as author, outlet, or engagement, analyzing a diverse set of article types allowed us to examine how different styles of vaccine-related content measured against traditional journalistic quality standards. Our results and learnings from this exploration are below.

Results

Each article was given a score between one and five, with a high score of five indicating that the article contained original reporting, meaningful analysis and good sourcing (more detail is provided in the "Overtone Scoring Definitions" section). The dataset was organized to ensure that there was a representative sample of articles in each scoring category.

Overtone algorithm rates text according to the presence or lack of journalistic signals. A key aspect of Overtone's scoring, especially regarding content from non-journalistic sources, is that while a low score may indicate low *journalistic* quality, it does not necessarily indicate low *information* quality. For example, some pieces of purely informational content from Vaccine Safety Net (VSN) member websites received a score of one. They did not receive a low score because the information contained in the article was incorrect, but rather, because the article lacked the signals of journalistic quality. In some cases, the VSN article might not have quoted its sources in a traditional journalistic style,[1] or did not include citations from additional independent sources (because the information was coming directly from an authoritative source). In other cases, the article may have resembled a press release, in language style or form of explanation.[2] These examples reiterate that in our dataset, low scores do not always

[1] "Cancer, Misinformation, and Education: Why Canada Is Far behind Rwanda's Vaccine Coverage," I Boost Immunity, April 1, 2021, https://iboostimmunity.com/content/cancer-misinformation-and-education-why-canada-far-behind-rwandas-vaccine-coverage.
[2] Communicating with Families and Promoting Vaccine Confidence." 2021. American Academy of Pediatrics. July 30, 2021. http://www.aap.org/en/patient-care/immunizations/communicating-with-families-and-promoting-vaccine-confidence/

indicate an unreliable or poor article. Low-scoring articles may provide good, truthful coverage, but receive a low score simply because they do not contain the amount of explanation or reporting that higher-scoring articles do.

A second aspect of scoring deals with the amount of journalistic effort put into an article. The treatment of aggregation in our dataset illustrates this point. Aggregation within an article occurs when a journalist takes information from another source or reporter, and uses it as the main source of information in an article. For example, this NY Post article that pulls from an article at Fox News as its source, then later pulls information directly from a Washington Post article, receives a score of two in our dataset.[3] Articles in this "aggregation" group will often receive a low score, because actions such as copying and pasting from another website represent low journalistic effort.

A third notable aspect of our dataset is the clear focus on content, not source, evaluation. For example, an outlet can produce multiple types of stories that receive different scores. A short CNN article covering an executive order on vaccine mandates is scored as a one, because it is essentially reproducing a press release, while a more in-depth CNN article on vaccine skepticism in Romania that provides strong ground reporting and nuanced analysis is scored as a five[4] [5]. This agnosticism toward source evaluation is also particularly evident when newswires are used. The algorithm cannot distinguish between a newswire on one website and the same newswire copy on another site. A well-reported piece of content from the Associated Press will be awarded with an accordingly strong quality score, no matter which masthead the content may appear under. In a similar vein, a vaccine-related article from the Daily Mail may receive a one for its lack of journalistic signals, but when the Daily Mail uses a newswire from the Associated Press, it can receive a four.[6]

Ultimately, it may be that this dataset provides its most useful insight not from low scores, but from the content scored as a four or five. This content is more likely to represent a high-quality journalistic and deeply informational vaccine article - one that uses specialized, reliable sources and subject matter experts, includes nuance and context, and appropriately handles scientific uncertainties.

Learnings

We were able to glean some valuable insights from the results of the article analysis in this dataset. First, it provides some initial guidance on which type of vaccine-related news articles might be worth recommending. Many datasets in the field of quality and credibility rate articles with binary "reliable" or "unreliable" veracity labels, generally at the source level. But, there can

[3] Natalie O'Neill, "Fauci insists there's no need to cancel Christmas during pandemic ," *New York Post*, December 17 2020, https://nypost.com/2020/12/17/fauci-says-canceling-christmas-during-pandemic-is-nonsense/
[4] Paul LeBlanc, "Texas governor bans Covid-19 vaccine mandates by any employer in state," *CNN*, October 12 2021, https://edition.cnn.com/2021/10/11/politics/texas-vaccine-mandate-greg-abbott/index.html
[5] Christiana Moiescu, "Bodies pile up outside hospital morgue as Romania struggles with fourth wave of Covid," *CNN,* November 23 2021, https://edition.cnn.com/2021/11/22/europe/romania-covid-19-vaccine-skepticism-intl-cmd/index.html
[6] Caitlyn Becker, " Ramona Singer supports Tweet comparing COVID vaccine mandates to 'Nazi territory' ... following RHONY reunion being cancelled after Eboni K. Williams 'filed a formal complaint against' her over a 'racist' comment," *Daily Mail,* November 2 2021, https://www.dailymail.co.uk/tvshowbiz/article-10158219/Ramona-Singer-supports-Tweet-comparing-COVID-vaccine-mandates-Nazi-territory.html.

be logical fallacies, incorrect representation of claims, and other low-quality signals even among highly reputable news sources. Our dataset shows the spectrum of journalism quality, and by doing so, highlights the type of in-depth or investigative reporting that represents higher quality information.

Second, it affirms that there are different considerations in quality evaluation for general journalism and quality evaluation for scientific or health communications. In our dataset, strictly informative pieces from authoritative sources, such as content from VSN members, do not perform as well with a journalistic-based algorithm because they are being evaluated against traditional news articles. This raises the point that perhaps, in some forms of journalism, the source or outlet is a worthy metric to consider when evaluating quality.

And finally, it is important to note that this software, and textual content analysis in particular, is an evolving space. If the articles being evaluated are significantly different enough from the data that the model was trained on, then the scoring is going to be more variable in non-traditional articles than in more standard articles. An article can also be mis-scraped, which can lead to a lower or higher score than deserved. But despite the analytical and data challenges, the need for automated quality assessment continues to grow. Thus, this dataset, and its application of quality assessment to vaccine-related content, aims to contribute to that research.

About ARTT
The Analysis and Response Toolkit for Trust (ARTT) project is focused on helping people engage in trust-building ways when discussing vaccine efficacy and other topics online.

About Overtone
Overtone has built a Natural Language Processing algorithm that finds and sorts online content by its intrinsic qualities, rather than clicks or shares. Their AI assesses texts for journalistic signals that demonstrate human effort.

For questions, please contact artt@hackshackers.com