



ad fontes media

Content Analysis Methodology Paper

Updated December 2024

Authors: Vanessa Otero, Jeff Von Wald, Erin Fox-Ramirez, and Beth Heldebrandt

I. Overview

II. Core Taxonomy

1. Framework
2. Definitions

III. Analysts

1. Education and Qualifications
2. Political Leanings
3. Training

IV. Process of Analysis

1. Content Selection
2. Analysis

V. Data Visualization

VI. Data Quality

VII. Machine Learning/AI

VIII. Continuous Improvement

I. Overview

The founder and CEO of Ad Fontes Media (AFM), Vanessa Otero, created the first Media Bias Chart® in October of 2016 as a hobby for the purpose of creating a visual

tool to discuss the news with friends and family. Otero has a B.A. in English from UCLA and a J.D. from the University of Denver Sturm College of Law and practiced patent law for a total of six years. She founded Ad Fontes Media in 2018 and left her law practice in 2020 to run the company full-time.

Today, Ad Fontes rates thousands of news sources for reliability and bias, and its content methodology is widely recognized as one of the most robust and comprehensive systems available for measuring news content. This white paper describes, in detail, the process by which Ad Fontes Media currently produces the various iterations of the Media Bias Chart®. Aspects of the systems and methods described herein are patent pending.

Otero created the taxonomy of the chart and analyzed the initial set of news sources herself. However, as the chart grew in popularity, she sought to improve the methodology, make the process more data-driven, and mitigate her own biases. To do so, she recruited teams of politically diverse analysts and trained them in the methodology. Over time, this process evolved into Ad Fontes Media's current method of multi-analyst content analysis ratings.

The methodology has evolved with input from various commentators and industry experts, including AFM Advisor, longtime journalist and journalism professor [Wally Dean](#), who has worked at the Pew Center for Civic Journalism, the Project for Excellence in Journalism, and the University of Missouri at points during his illustrious career. Dean co-authored [We Interrupt this Newscast](#), a book detailing one of the largest content analysis studies ever done.

History

AFM conducted its first multi-analyst content ratings in 2019 with a group of over twenty analysts. Over the next year, nine analysts involved in the research stayed on at AFM to rate several dozen articles every month as well as to add new sources and update previously existing ones. Since then, AFM has trained a total of 117 analysts to continuously rate news articles, digital video, TV programs and podcasts. As of Dec. 13, 2024, the AFM dataset includes 82,300 multi-analyst ratings of articles and episodes from 2,600 web/print sources, 766 podcasts and 700 TV shows, and the list is growing every day.

Since the release of the original Media Bias Chart® in 2016, millions of observers have found this system of classification to be useful. AFM has also received criticisms and suggestions from social scientists, data analysts, statisticians and news organizations on how this work could be improved, and the feedback is welcome. Some news organizations have found our data useful enough that they've asked us to audit their content to assess political bias or factual shortcomings. The news ecosystem is

continuously shifting, and we recognize that there will be a need for the methodology to evolve as well.

Some observers take issue with aspects of the taxonomy itself; for example, there are objections that a left-right axis does not capture the full range of possible political positions. Others object that placing news sources in an infographic format, in and of itself, strips complex concepts of reliability and bias of their nuance. Others object that the placement of sources on both the left and right reflect a false balance.

However, the media landscape is vast, and issues of reliability and bias are inherently complex. The two-dimensional framework and visual presentation make the data easily accessible to a wide audience. Our interactive version of the chart, detail-rich database, and extensive methodology documentation provide deeper levels of nuance for those inclined to dig deeper. Based upon the proven utility of this taxonomy and the increasing demand for the underlying data, the chart remains a helpful visual representation of the data, and our processes and methodologies undergo improvement and refinement along the way.

Why we rate the news using Content Analysis

“Ad Fontes” is a Latin term for “to the source.” We chose that name because it reflects the fact that we focus on the content – what’s in the news source itself – to make determinations about reliability and bias. There are other methods of assessing the trustworthiness and bias of a news outlet, the primary one of which is polling. However, polling has certain limitations. One is that asking a polled group about their perceptions of trustworthiness or bias of an outlet is highly subjective and inherently limited to what each person can recall from their past experiences with such outlets. Another limitation is that it is difficult to poll people about smaller and more obscure news sources because fewer people are familiar with them.

Content analysis allows us to mitigate subjectivity and apply the analysis to all news and information sources – even those that are small and obscure.

Content Analysis is a widely used methodology in communication research and is quite common in media studies because it has the following advantages:

Objective - Content Analysis is a systematic and structured approach to analyzing the content of a news story. It's observable and measurable in terms of looking at words, themes, sources and framing. CA also reduces subjective interpretations and personal bias.

Reproducible - Content Analysis allows for replication by other analysts and for comparisons between analysts, which together increases the validity of the results.

Quantitative - Content Analysis generates quantitative data, a numerical representation that allows for statistical analysis and comparisons that can help

identify patterns within the corpus and across corpora.

Comprehensive - Content Analysis is comprehensive by including multiple elements in the analysis of news stories, including language, headlines and visuals.

Large-Scale Analysis - We can apply Content Analysis to large samples of news stories, identifying broader trends and patterns. This is especially useful when studying media across different outlets and over time.

The Media Bias Chart is a Taxonomy and a Methodology.

Although the Media Bias Chart® strikes many observers as an intuitive infographic, it is really the combination of a defined taxonomy (a system of classification) and a methodology (a repeatable process) for placing news and informational content within the taxonomy.

AFM currently provides in-depth discussions of the taxonomy and [methodology](#) in various public-facing videos and webinars, including an annual free teacher training. Several videos from these training sessions are linked throughout this paper to provide additional detail on the rating process.

II. Taxonomy

A. Framework

Our taxonomy is a two-dimensional framework for categorizing the reliability and bias of content, shows and sources.

The horizontal axis (political bias, left to right) is divided into nine categories, four of which represent the spectrum on the left, four of which represent the spectrum on the right, and one in the middle. Each category spans 12 units of rating, so the total numerical scale goes from -42 on the left to +42 on the right. These values are somewhat arbitrary, though there are some good reasons for them, including that they 1. allow for at least seven categories of bias, 2. allow for more nuanced distinction between degrees of bias within a category (allowing analysts to categorize something as *just a bit* more biased than something else), and 3. correspond well to visual displays on a computer screen or a poster.

Bias scores are on a scale of -42 to + 42, with higher negative scores leaning more to the left, higher positive scores leaning more to the right, and scores closer to zero being centrist, minimally biased and/or balanced.

It is important to note that “middle” doesn’t necessarily mean “best” on the Media Bias Chart. News sources can land in the middle section for bias for at least three

reasons: 1. the content is biased toward a centrist or “neither-side” position, or 2. the content presents a balance of biased political positions, which can include a balance of far-left and far-right arguments. Content described in examples 1 and 2 can score low on the reliability scale while being in the middle of the bias scale. Content can also land in the middle because it is 3. minimally biased; that is, delivering factual information as straightforwardly as possible, with minimal characterizations. This type of content does have a high correlation with high reliability scores, so there is a high density of fact-reporting content that has bias scores near 0.

The vertical axis (overall reliability, top to bottom) is divided into eight categories, each spanning eight rating units, for a total numerical scale of 0 to 64. Again, these are somewhat arbitrary, but the eight categories provide sufficient levels of classification of the types of news sources we are rating and sufficient distinction within the categories. Reliability scores are on a scale of 0-64, with source reliability being higher as scores go up.

For a more detailed background on why our scoring system is from 0-64 and -42 to +42, see [this document](#).

Overall source ratings are composite weighted ratings of the individual article and show episode scores. Individual content pieces that have reliability scores below 24 and bias scores higher than +/- 6 are weighted more heavily.

There are several sub-factors our analysts take into account when considering the reliability and bias of an article, episode, or other content. The main ones for *Reliability* are defined metrics we call “Expression,” “Veracity” and “Headline/Graphic,” and the main ones for *Bias* we call “Political Position,” “Language” and “Comparison.”

Reliability sub-factor descriptions:

The “**Expression**” sub-factor accounts for how an article is expressed – as fact, analysis or opinion.

The “**Veracity**” sub-factor accounts for the truth or falsity of explicit and implicit claims. This is where we incorporate fact-checking as part of the reliability metric. The “Veracity” factor is of particular importance in the reliability score. Our analysts use a veracity-checking methodology that incorporates best practices of fact-checking, such as lateral reading and consulting primary sources, but which is designed to be broad enough to cover claims that are not fact-checkable and quick enough to make an evaluation on every article. For more information on our veracity evaluation methodology, see [this video](#).

The “Headline/Graphic” sub-factor accounts for whether the headline and graphics match the content of the story and the impact of the headline and graphic in relation to the claims in the story.

Bias sub-factor descriptions:

The “**Political Position**” sub-factor accounts for advocacy of political positions, actions or politicians in an article or episode.

The “**Language**” sub-factor accounts for terminology used to characterize political issues and opponents.

The “**Comparison**” sub-factor accounts for bias due to topic selection or omission.

B. Definitions

The horizontal (or “**Bias**”) categories are defined by the policy positions of current U.S. elected officials. For more on why, [see this methodology video](#). This video also discusses how the U.S. left-right spectrum shifts over time. This concept is related to, but distinct from a concept known as the Overton window. Because we rate media for its left-to-right bias, we need a baseline to which we can compare the media rating. We define areas of the horizontal axis, particularly with regards to the “political position” subfactor, as follows:

1. The line between “Most Extreme Left/Right and Hyper-Partisan Left/Right” is defined by the policy positions of the most extreme *elected* officials significantly relevant to the scope of the issue being considered.
2. The line between “Hyper-partisan Left/Right and Strong Left/Right” is defined by the current policy positions and actions of *median* leaders of the major left and right parties.
3. The “Strong Left/Right” and “Skews Left/Right” subcategories mark the degree to which a policy position is closer to or farther away from the “Hyper-partisan” and “Middle” categories.
4. The “Middle or Balanced Bias” category is labeled as such because content can fall in the middle for one or several reasons. We don’t label it as “neutral” or “unbiased” because all content has some kind of bias. As described previously, content can land in the middle because it has a centrist or neither-side bias; or because it is balanced, showing two or more biased sides of an issue in similar degrees; or because it is minimally biased, stating facts as straightforwardly as possible.

An article, episode or source placing near the midpoint on the horizontal axis may land there for any of these reasons; thus the position does not necessarily represent “neutrality.” Nor is the midpoint on the horizontal axis intended to imply that the position is best or most valid.

The vertical axis, labeled on our chart as **“News Value and Reliability,”** but often referred to simply as “Reliability,” represents a continuum measuring how much a news or information source may generally be *relied* upon to present new information that is dense with facts, true, impacts people’s lives, and would be difficult for people to find on their own, as follows:

1. Analysts score content they deem to be primarily fact reporting between 48 and 64, with the highest scores reserved for encouraging the hard (and socially essential) work of original fact reporting that is subsequently corroborated by additional sources.
2. Content that includes analysis scores between 32 and 48, with the higher scores in this range reserved for analysis that is supported by well-argued fact reporting. In terms of “reliability,” the taxonomy places opinion (24-32) below analysis. However, as with analysis, opinion that is well-argued and defended based with facts also scores higher within the category.
3. Content scoring below 24 generally has a reliability problem. When it scores between 16 and 24, very likely an important part of the story was omitted. It is likely (and literally) a “partial” story representing – at least in that sense – an “unfair” attempt at persuasion. Content scoring below 16 has been determined by our analysts to be misleading or downright false, at least based on the best evidence presented to date.

It should be noted that our taxonomy and methodology constitute a rubric used to *describe* content on both the horizontal (“bias”) and vertical (“reliability”) axes. Bias scores are *descriptive* in relation to the current politics of the country as a whole. They are not intended to rate the moral quality of a position; nor are they measured against a timeless or universal norm. Reliability scores are similarly *descriptive*, though veracity is one of the metrics considered on the vertical axis. While veracity is part of what we consider when rating the reliability of content, there is a categorical difference between the “rightness” or “wrongness” of content expressing an opinion and the “truth” or “falsehood” of content stated as fact. Moreover, there are limits to human knowledge, and our methodology considers “likelihood of veracity” to be more accurate than a “true/false” toggle when considering the accuracy of content presented as fact.

The overall source rating is a result of a weighted average, algorithmic translation of article raw scores. Low-quality and highly-biased content weight the overall source down and outward. The exact weighting algorithm is not included here because it is proprietary, but generally, lower reliability and more biased content is weighted more heavily.

III. Analysts

Since October 2020, Ad Fontes Media has employed a team of analysts to rate news content on an ongoing basis. At the time of this 2024 revision, we have 45 active professional analysts.

Currently, our analyst application process requires the following:

- Submission of a professional resume, CV or similar written summary of qualifications.
- Completion of an online application enabling further assessment of qualifications.
- A self-reported classification of their political leanings. Each analyst submitted a spreadsheet about their political views overall and per listed political topic. The “political position assessment” can be viewed on our site on the [analyst application](#) page.

Submission of basic demographic information is optional but helpful in maintaining an analyst team that is relatively representative of the country as a whole. Our [“Team” page](#) has statistics about the demographics of our analyst team as compared to the U.S. population.

A. Education and Qualifications

Our current qualification expectations for new applicants are as follows:

- Lives in the United States, and is politically/civically engaged
- Is familiar with a range of news sources
- Is familiar with party platforms and government systems in the U.S.
- Is willing to divulge political leanings internally as required by our approach to analysis
- Demonstrates excellent reading comprehension skills
- Demonstrates excellent analytical skills
- Demonstrates ability to engage in sometimes difficult conversations, including on sensitive issues
- Demonstrates ability to see issues from multiple perspectives while also respectfully expressing a dissenting perspective when applicable
- Demonstrates a passionate interest in news media and contemporary U.S. politics
- Demonstrates a desire to make a positive difference
- Has earned a Bachelor’s degree or equivalent. Currently enrolled college/university students may be considered if they have completed at least two years.
- Preferred applicants have advanced degrees, or a highly relevant undergraduate degree, in Media, Journalism, Political Science, Linguistics, History, Sociology, Philosophy, or other field requiring strong skills in analyzing information content.

- Helps contribute to the range of special subject expertise within our team
- Demonstrates familiarity with identifying bias and reliability in news sources
- Demonstrates interest in Ad Fontes Media and our mission

The extent to which applicants demonstrate the qualifications above is assessed by a politically balanced team of application reviewers using a shared rubric to identify the most qualified applicants.

All of our current analysts hold at least a bachelor's degree, and most have completed at least one graduate degree program. Approximately one-third have completed a doctoral degree program or are current doctoral students.

While education is an important qualification, a number of other factors are considered as well, particularly familiarity with U.S. politics and the ability to engage in rigorous critical reflection on written and spoken content. Analysts come from a wide range of professional backgrounds – including federal service, law, and management – the backgrounds represented most within the team are journalists, teachers, and librarians.

B. Political Leanings

Because analysts use a granular methodology and are looking for very specific factors while scoring, their scores of each piece of content are generally quite close regardless of the analyst's political bias. However, to mitigate the effect of any one analyst's bias, since 2019 each piece of content we rate has been rated by an equal number of analysts (i.e., three or more) who identify as left-leaning, center-leaning, and right-leaning politically.

To arrive at the classification of the analysts, we lean heavily on their own sense of political identity, along with a self-assessment that asks analysts for their political positions on the following categories:

1. Abortion-related policy
2. Race-related policy
3. Campaign finance
4. Climate-related policy
5. Criminal justice reform
6. Defense/military budget
7. Subsidized food and housing
8. Gun-related policy
9. Higher education policy
10. Immigration
11. International affairs
12. K-12 education policy
13. LGBTQ-related policies

14. Marijuana policy
15. Private/public health care funding
16. Regulation of corporations
17. Social security
18. Tax-related policies

For each of the issues above, we request that each analyst identify their perspective as:

- “Decidedly to the left”
- “Moderately to the left”
- “Centrist or undecided”
- “Moderately to the right”
- “Decidedly to the right”

For each issue in which the analyst identifies their perspective as “decidedly to the left,” they score “-2,” for each “moderately to the left,” they score “-1,” for each “centrist or undecided,” they score “0,” and so on.

Analysts scoring more than 10 points from 0 are initially categorized as left-leaning or right-leaning. Analysts scoring fewer than 4 points from 0 are initially categorized as centrist. Analysts falling between 4 and 10 points from 0 are considered on a case-by-case basis, with the analyst’s political identity considered most heavily.

In addition to this internally developed assessment tool, we also use two popular and publicly available political affiliation assessment tools – the [Political Compass](#) and Pew’s [Political Typology Quiz](#). We use multiple assessment tools because no one individual tool captures all the nuance of political identity, and what constitutes left, right, and center changes over time. We also implement a peer review assessment, asking analysts to place other analysts into the political category to which they feel they belong. This provides an additional check on any one person’s political bias, which can change over time, and ensures that our content is rated by analysts representing three different political viewpoints. By using a variety of assessments and updating our own assessment, we can mitigate the shortcomings of any one particular assessment.

While individuals’ political outlooks are generally quite complex and are often varied across issues, the analysts are generally able to identify their own perspective on these issues quickly using the framework above. To do so assumes a level of familiarity with U.S. politics, which is assessed during the application process. When combined with the practice of having each piece of content analyzed by an equal number of left-, center-, and right-leaning analysts, we have found that this system of classifying analysts helps mitigate the bias of any one analyst.

C. Training

Before joining the analyst team, each analyst trainee reads an article overviewing each step within Ad Fontes Media's eight-step core analysis methodology. The eight steps are made up of Veracity, Expression, Headline/Graphics, and Overall Reliability for the "Reliability" (vertical) metric, along with Political Position, Language, Comparison, and Overall Bias for the "Bias" (horizontal) metric. For each of these eight metrics, analysts also attend a 60-minute presentation on the same topic.

As they complete the assignments above, trainees also practice rating content using the methodology as a rubric. They also observe live analysis shifts where experienced analysts consider content together. Trainee ratings are observed; however, during this time, trainee scores are not included in the data used in our overall source and content ratings.

Upon successful completion of 30-35 hours of training described above, and once significant outlier scores are rare, trainees enter a probationary period where they score articles along with two experienced analysts. At this point, trainees' scores are included in source and content ratings, and any outlier scores are managed as described below.

All analysts attend ongoing training, which includes occasional fine-tuning to the methodology and awareness of the shifting meaning of categories such as "left" and "right" when applied to specific issues over time.

IV. Process of Analysis

A. Content Selection

To date, we have fully rated 4,000 sources, including web/print, podcast, and television/video formats. Members of our team use reach data, source lists, and user requests in order to select sources to be rated. While all sources gain additional article and episode scores over time, some sources have many more data points than the minimum. No source or show is considered to be "fully rated" until our team has rated a minimum of 15 articles for web/print content or three complete episodes of podcast or television content. To date, our team has scored over 82,300 articles and episodes to date in order to arrive at approximately 4,000 fully rated sources, and analysis is ongoing with several shifts of live analysis running daily.

Upon selecting a source to be rated, we select a sample of articles or episodes for analysis, and the scores of these individual content pieces inform the overall source score.

Articles are currently selected manually based on their "prominence," as determined by page placement, size of print headline, or when available, based on reach. Prominence functions partly as a proxy for reach and is an important part of our

methodology because many publishers feature highly opinionated or biased content to drive engagement, even if most content they publish is more fact-based and neutral. Public perceptions of bias of large publishers are often driven by the extensive reach of lower-reliability, highly biased content.

For TV networks, content is similarly selected based on reach and its prominence in terms of when it is scheduled to air. For podcasts and TV shows, sample episodes are selected based on their representation of the show overall.

For some sources, current ratings are based on our minimum sample size, which may be small. However, if within this minimum sample we notice wide variation in reliability or bias scores, we increase our sample size until the overall score stabilizes within a range. This is one way we ensure our sample articles and shows are sufficiently representative of their respective sources. However, the larger the sample size, the more precise the overall score becomes over time, so we strive to increase the sample size of all sources on an ongoing basis.

We rate all types of content, including those labeled analysis or opinion by the news source. Not all news sources label their opinion content as such, so regardless of how it is labeled by the news source, we make our own methodology determinations on whether to classify articles as analysis or opinion on the appropriate places on the chart. For more detail on why we do this, see [this blog post](#).

The content rating period for each rated news source is performed over multiple weeks in order to capture sample content over several news cycles. Sources that have appeared on our Media Bias Chart for longer have content from more extended periods of time.

Often, our sample sets of articles and shows are pulled from sites on the same day, meaning that they were from the same news cycle. Doing so allows analysts to incorporate evaluations of bias by omission and bias by topic selection.

We update all sources periodically by adding new content. Because we have so many news sources, and because the most popular sources are important to the public, we generally update the most popular sources more frequently and less popular sources less frequently. For example, we update a tier of the top 50 sources with at least five new manually rated articles each month, and the top 1,000 with at least 12 new articles per year. We strive to balance rating new sources and updating existing ones.

B. Analysis

Each individual article and episode is rated by at least three human analysts with balanced right, left, and center self-reported political viewpoints. That is, at least one person who has rated the article self-identifies as being right-leaning, one as center-, and one as left-leaning.

The main principle of Ad Fontes (which means “to the source” in Latin) is that we analyze content. We look as closely as possible at individual articles, shows and stories, and analyze what we are looking at and hearing: pictures, headlines, and most importantly, sentences and words.

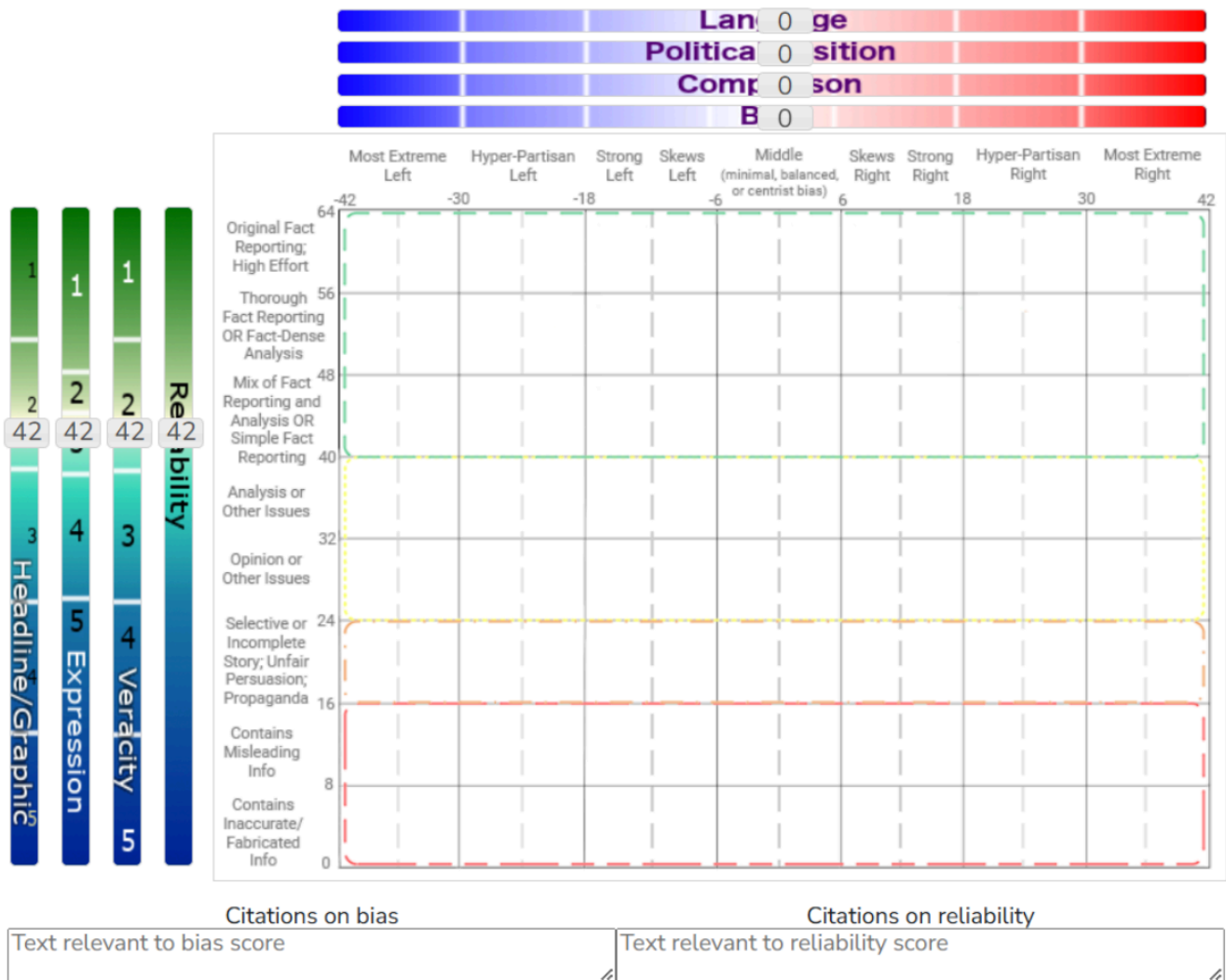
Since 2020, we have rated most content in three-person synchronous shifts because this “live” process requires each analyst to justify their score when needed, aids in exposing analysts to multiple perspectives, and allows analysts to point out aspects of the content that may have been missed by a single person in the group.

Analysts meet in two-hour shifts and go through pre-assigned articles and episodes together on Zoom. One of the analysts is assigned to be the facilitator for the shift. For articles, each analyst reads the article on their own, scores it, and then the facilitator displays all of the analysts’ scores together. For episodes, the facilitator plays the episode for all analysts to listen to/watch together, pausing periodically for discussion and note-taking, and then each analyst scores it on their own. The facilitator displays all analyst scores for the group. If all scores are within an eight-point range for both reliability and bias, the three scores are averaged to make up the overall score for the article or episode. If the scores are not within range, the facilitator leads a discussion between all analysts to explain what they considered in their scores. Analysts may then adjust their scores if they find the reasoning of others to be persuasive such that the scores do fall within range. If the analysts cannot get their scores within range, the article or episode is sent to a second panel of analysts.

Occasionally, content is rated asynchronously by experienced analysts for logistical reasons. However, when analysis is done asynchronously, our commitment to a politically balanced multi-analyst approach to each piece of content remains – it still receives a rating from one left-leaning analyst, one right-leaning, and one center.

Analysts enter their scores into our proprietary software platform known as “CART,” which stands for Content Analysts Rating Tool. Our analysis operations team enters articles and episodes to be rated into this system, which automatically parses article information such as the headline, author(s), all the article text, and some statistics on the text, including word count, parts of speech, number of questions, and certain ratios between parts of speech.

Analysts use a blank Media Bias Chart® interface to enter their overall reliability and bias scores. The interface has sliders for assigning scores to each of the reliability and bias subfactors as well, as shown below:



In addition to the scores, analysts are required to enter notes indicating reasoning for the reliability and bias scores, with citations to the text and language in the article or episode.

V. Data Visualization

The easiest way to see the resulting ratings for each article and show is, of course, on the [Interactive Media Bias Chart](#)®. By clicking on the name of a source in the search box, you can see a scatter plot of each article or episode rated for that source. The center of the logo of the news source is placed where its overall score is, which is a weighted average of its individual content piece score.

Each dot represents an individually rated content piece, and you can click on the dot to view the article or episode that was rated.

Close observers of the Interactive Media Bias Chart® will notice that, particularly for low-scoring sources, the overall source scores appear to be lower than what would be

expected from a straight average. As previously mentioned, this is because in our overall source-rating methodology, we weight extremely low reliability and extremely high bias content scores very heavily.

The reason for weighting is this: the lowest rows of the chart indicate the presence of content that is very unreliable, including selective or incomplete stories, unfair persuasion, propaganda, misleading information, inaccurate, and even fabricated information (these are listed in order of egregiousness). Therefore, it is unacceptable for reputable news sources to include this type of content, even if it is infrequent or not the majority of the content. A source that has even 5% inaccurate or fabricated information is highly unreliable. A source that “only” publishes misleading or inaccurate content 33% of the time is terrible. In our system, they do not get credit for the 67% of stories that are merely opinion, but factually accurate.

A straight average, in such cases, would result in a higher overall source score – one that is inconsistent with the judgment of most savvy news consumers. Therefore, article scores of less than 24 for reliability are weighted very heavily. The weighting increases the lower the rated content falls under 24.

We also rate bias scores more heavily the farther the scores are away from zero. This results in sources with left- or right-leaning opinion content mixed with neutral/balanced content skewing more overall toward the bias of their opinion content. For example, The New York Times and Wall Street Journal skew left and right, respectively, due in large part to their opinion section content.

All other content scores for sources are straight-averaged. For example, if a news source only has a mix of “fact reporting,” “complex analysis,” “analysis” and “opinion” articles (no articles below 24), those would be straight averaged. As shown, our taxonomy rewards high percentages of fact reporting and complex analysis in sources and slightly down-ranks them for high percentages of opinion content (via straight averages). It does not punish a source for opinion content, because opinion content does have a useful place in our information ecosystem. However, our system does punish unfair opinion and worse content – that which we view as the most polarizing “junk news.”

VI. Data Quality

We have implemented several processes for continuous improvement of our data around the following areas:

1. Inter-rater reliability – Inter-rater reliability is a measure of consistency used to evaluate the extent to which different analysts agree in their rating scores. Training currently involves having all analysts rate certain content to capture this metric and provide additional training and feedback thereupon.

2. Intra-rater reliability -- Intra-rater reliability refers to the consistency of ratings or measurements made by the same rater or observer on multiple occasions. We collect statistics on individual analysts' average scores across all content they've rated and use internal calculations to identify tendencies.

3. Sampling – Over time, we have been able to increase the base sample size of “fully rated” source samples, and the growth of our operations will naturally result in continued increases in all manually rated samples.

4. Sunsetting – To keep our database current and in line with what news sources have published most recently, we have begun phasing out old scores that are more than five years old. We will keep these scores for archival and research purposes. We are implementing the capability to search for changes in a source's overall reliability and bias scores over shorter time windows.

VII. Machine Learning/AI

In August of 2023, Ad Fontes announced its capability to rate articles using machine learning. Ad Fontes developed a proprietary machine learning model trained on its data set of over 70,000 human-labeled pieces of content at the time. This machine learning model predicts scores of new articles from news sources that have been previously rated by Ad Fontes Media. The model implements existing natural language processing and machine learning techniques as well as numerous custom features derived from text signals we have found particularly useful for predictions regarding reliability and bias.

As of the publication date of this White Paper, we currently rate over 120,000 news articles per day in substantially real-time with our ML model, which we provide for commercial use to stakeholders in the media ecosystem who desire reliability and bias scores for making decisions at the page level.

Our primary measure of the accuracy of our ML model ratings is the Mean Absolute Error (MAE), which is the average of how many points on our scale the machine rating is from our human analyst rating. We capture and monitor this on a continuous basis because every day, our human analysts continue to rate articles, and we run those human-rated articles through our ML model for scoring and compare them. Currently, our mean absolute error for both reliability and bias is approximately four points, which means that on a scale of 0-64 for reliability, our model scores articles on average within four points of our human ratings, and on a scale of -42 to +42 for bias, our model scores articles on average within four points of our human ratings. We consider the MAE of 4 points to be quite accurate, given that the standard deviation between human analysts is approximately three points.

VIII. Continuous Improvement

Ad Fontes Media is committed to continuous improvement of its methodology. As the news and information landscape continues to evolve, and as we recognize better ways to measure content for reliability and bias as objectively as possible, we will implement such improvements. We welcome all suggestions on how to do so. For more information, please contact info@adfontesmedia.com.