Introduction:

To demonstrate the differences in media coverage between our four groups, we performed sentiment analysis upon 511 case articles of our 515 cases. We excluded ten case articles, all of which were in the Non-Control Non-AMI group, for a variety of reasons that we will explain shortly. When citing these and other cases, we will use their standardized names as documented on GitHub.

Methods:

Initially, we attempted to scrape the text of every article or other link using the Trafilatura package in Python because its implementation was exceptionally straightforward, good at eliminating textual noise, and almost universally applicable to different HTML formats, which was in stark contrast to other packages like Beautiful Soup that we also attempted to configure. However, even the incredibly versatile Trafilatura had its drawbacks, occasionally scraping some pages as blanks or even failing to scrape them entirely. In such instances, we either manually scraped the articles or removed their links if we deemed their content unsuitable for our analysis.

When it comes to the former, we scraped the following four case articles manually, mimicking the compact Trafilatura format as closely as possible:

- Sarah Hyland; Matt Prokop Case
- Jessica Canseco; Jose Canseco Case
- Oksana Grigorieva; Mel Gibson Case
- Martine Roy, Marylena Sicari, Guylaine Courcelles; Gilbert Rozon Case

As for the latter, we omitted the following three case links because they did not lead to proper articles:

- John Stamos; Finland Woman Case
- Driz Family; Matthew, Anne Greene Case
- Michael Ovitz Case

Note that this accounts for three of the ten excluded case articles.

In one observed glitch, Trafilatura neglected to scrape parts of a US Magazine article for the John and Jane Does; Jeff Garlin Case that were beyond the sentence, “Scroll below for everything to know about *The Goldbergs’* investigation into Garlin’s show future.” For completeness, we manually scraped the omitted text and added it to the included portion.

Further, although they are articles in a literal sense, we also excluded any and all Wikipedia pages regardless of how well they were scraped since they aren’t in the same category as news media, the true target of our analysis. Nevertheless, we will still give a full accounting of our omissions. Firstly, the following two Wikipedia articles were scraped as blanks:

- Johnny Depp; Amber Heard Case
- Dylan Farrow, Mia Farrow; Woody Allen Case

Secondly, two other Wikipedia pages were scraped properly, the second of which corresponded to three cases and is therefore linked once for each:

- Amanda Segel; Bob Weinstein Case
- Marc Collins-Rector; Minors Case
  - Minors; Chad Shackley Case
  - Brock Pierce; Minors Case

Altogether, we eliminated six case articles associated with Wikipedia, which also contributes to the ten excluded case articles.

Thirdly, since we used the same ABC article for the following two cases, we had to omit one instance thereof:

- Stan Lee; Maria Carballo Case
- Maria Carballo; Max Anderson Case

When combining the three improper links, six Wikipedia pages, and this ABC article, we arrive at the original figure of ten omitted case articles that we initially outlined. At this point, one might be puzzled, as the difference between our original 515 cases and our analyzed 511 articles is not ten but four. If we removed ten articles, should we not have only 505 remaining? In reality, this discrepancy is spurious, as it results from the conflation of cases and articles. While we did indeed start with 515 cases, we also began with 521 articles because the following six cases each had two associated articles or other links, all of which remain in our sentiment analysis besides, as we noted earlier, the Wikipedia article for the Dylan Farrow, Mia Farrow; Woody Allen Case and the Instagram link for the Driz Family; Matthew, Anne Greene Case:

- Justin (Sane) Geever; Kristina Sarhadi Case (Second Link)
- Allison Mack; Jane Does Case (Second Link)
- Dylan Farrow, Mia Farrow; Woody Allen Case (Second Link)
- Tupac; Ayanna Jackson Case (Second Link)
- Kathryn Dennis Case (Second Link)
- Driz Family; Matthew, Anne Greene Case (Second Link)

In fairness, the Reddit post included as the main case link for the Tupac; Ayanna Jackson Case is not a news article either and should also be excluded. In a future study, researchers should make this and other revisions. But since this oversight affects just one data point, which is not necessarily an irrelevant one given that social media is just as if not more prone to bias than traditional media, we find it unlikely that it compromised the quality of our analysis.

Regardless, with our reductions complete, we were left with the following group article counts:

- Control AMI: 11
- Control Non-AMI: 88
- Non-Control AMI: 178
- Non-Control Non-AMI: 234 (from 244)

Next, we performed our sentiment analysis using the TextBlob library, as it was, like Trafilatura, fairly accessible in its implementation. We attempted to use the more sophisticated features of the Transformers library too, but the 512 token limit was ultimately not worth the effort of circumvention.

Using TextBlob, we generated both polarity and subjectivity scores for each article we analyzed. Then, to assess the distribution of these scores, we derived their means and population standard deviations within each group, each pair of like groups, and the reduced dataset as a whole. In the next section, we will present our results.

Analysis Results:

In summarizing both our polarity and subjectivity analyses, we have constructed two tables with each cell in mean ± standard deviation format and each figure rounded to five decimal places. Thus, in the following pages, any discrepancy between our derivations downstream from the tabular data and those one might reproduce independently are due to the precision lost by said rounding.

To start, let's examine our polarity table:

Group	Control	Non-Control	Total
AMI	0.03341 ± 0.06789	0.03718 ± 0.07846	0.03696 ± 0.07789
Non-AMI	0.05425 ± 0.07959	0.06937 ± 0.10107	0.06524 ± 0.09592
Total	0.05193 ± 0.07865	0.05546 ± 0.09336	0.05478 ± 0.09071

At first glance, since polarity scores in TextBlob have a range of [-1, 1], these results may seem rather unremarkable. After all, not only is every score mean slightly greater than zero, but the greatest among them is only 0.06937. Does this not make all our articles neutral if not marginally positive regardless of group?

While this is one way to interpret our results, it misses a critical insight, namely that the ratios of our mean polarity scores are often quite extreme. In fact, the ratio of the largest mean to the smallest (0.03341) is 2.07646, which starkly refutes the notion that our data is somehow uniform.

In that vein, notice that while the Control mean is only slightly less than the Non-Control mean, the AMI mean is approximately half (56.65602 percent) of the Non-AMI mean. Even though none of our polarities are negative, the AMI polarity scores being markedly less positive than the Non-AMI ones could suggest that AMI articles are far less positively disposed towards the victims than the Non-AMI ones. However, since TextBlob is analyzing not the treatment of individual entities but the text as a whole, this output is insufficient for drawing such conclusions.

Interestingly, recall that when we compared the Control and Non-Control groups, the trend was reversed, with the higher stakes of the Non-Control group coexisting with higher rather than lower polarity means. This is especially pronounced in the Non-AMI subset, as the Non-Control Non-AMI mean is 27.87693 percent greater than the Control Non-AMI mean. Thus, whatever factor correlates AMI involvement with lower polarity scores acts opposite that which correlates the severity of a case with higher ones.

That being said, it’s understandable that the group mean closest to the overall mean of 0.05478 is that of the Control Non-AMI group (0.05425). This indicates that removing story stakes and AMI activity eliminates in turn their positive and negative effects upon the magnitude of our polarity scores, respectively, leveling out to their mean.

On a side note, while all the standard deviations are also fairly small, notice that they too follow a trend. With the exception of the Control Non-AMI and Non-Control AMI standard deviations, which are only 0.00113 apart, the standard deviations rise with the size of each group. This is consistent with the notion that more data points allow for more variation within their dataset. Alternatively, since the standard deviations ascend as do the group means without exception, this could also just reflect the ease of deviating from a mean with a larger magnitude.

With these considerations in mind, let’s analyze this analogous table for the subjectivity scores by group:

Group	Control	Non-Control	Total
AMI	0.41895 ± 0.08397	0.40193 ± 0.08487	0.40292 ± 0.08491
Non-AMI	0.37837 ± 0.08130	0.41877 ± 0.10946	0.40773 ± 0.10410
Total	0.38288 ± 0.08259	0.41149 ± 0.09993	0.40595 ± 0.09747

Unlike in the polarity case, our figures are far from miniscule. Since TextBlob's subjectivity scores lie in [0, 1] with larger scores denoting greater subjectivity, all of our score means indicate moderate subjectivity with a leaning towards objectivity. Also unlike before, the ratios between means are far from extreme: the largest-to-smallest ratio is just 1.10725, a far cry from the 2.07646 figure we reported in the polarity case. Notably, the numerator of that ratio is Control AMI mean, which was the denominator in the polarity ratio, and the new denominator is the Control Non-AMI mean. Could this mean that increased subjectivity is associated with decreased or even less positive polarity? Unfortunately, since the Control AMI group has only 11 articles and the Non-Control Non-AMI group, which reigned supreme in polarity, is barely outpaced now, it would not be wise to make too much of this apparent inversion.

Even so, consistent with our initial conception, heighted AMI involvement seems to raise the subjectivity of an article, albeit not by much. However, when examining the pairs of like groups, we find that this trend does not stand up to scrutiny. While the Non-Control mean exceeds the Control mean as we would expect, the AMI mean is very slightly less (0.00481) than the Non-AMI mean. This might be confusing at first given that the Control AMI group mean surpasses the Control Non-AMI one by far more than the Non-Control AMI group mean exceeds the Non-Control Non-AMI one. But since the pairwise means are weighted by group size and the Control groups are altogether just 19.37378 percent of our reduced dataset, their subjectivity margins are overshadowed and ultimately reversed in the pooled statistics.

Note that this is essentially a less extreme iteration of the two trends we observed in our polarity analysis: Story stakes raise our scores while AMI activity somehow lowers them. Curiously, the standard deviations now ascend by neither group size nor score mean, although the Non-Control Non-AMI standard deviation maintains primacy at approximately 0.1 as before. In fact, it is the only group whose standard deviation maintains its ranking in both polarity and subjectivity analyses.

Conclusion:

As we clarified on several occasions, a key takeaway from this sentiment analysis is the tentative nature of its takeaways. Thus, we must proceed with caution in our conclusions accordingly. Nevertheless, based on our polarity and subjectivity distributions, we can assert the following with due confidence: AMI involvement is associated with lower polarity and subjectivity scores, whereas story stakes (i.e., those of the Non-Control group) are associated with a rise in said scores. Since all our polarity scores were positive, it remains unclear whether lower polarity scores are indicative of neutrality or a slight increase in negativity, which plausibly could be directed towards victims. A more advanced mode of sentiment analysis that isolates attitudes towards particular parties would mitigate such ambiguities and help unveil the harms of AMI reports. For now, let us appreciate the fact that this analysis demonstrates that all media, AMI or not, reports on the most consequential cases in the least objective and most polarized ways. In quantitatively illustrating this problem, we are one step closer to eliminating such reporting as well as the havoc it wreaks upon the victims in its unrelenting wake.