Verba Volant, Scripta Volant: Understanding Post-publication Title Changes in News Outlets

Xingzhi Guo
Stony Brook University
New York, USA
xingzguo@cs.stonybrook.edu

Brian Kondracki
Stony Brook University
New York, USA
bkondracki@cs.stonybrook.edu

Nick Nikiforakis
Stony Brook University
New York, USA
nick@cs.stonybrook.edu

Steven Skiena
Stony Brook University
New York, USA
skiena@cs.stonybrook.edu

ABSTRACT

Digital media (including websites and online social networks) facilitate the broadcasting of news via flexible and personalized channels. Unlike conventional newspapers which become "read-only" upon publication, online news sources are free to arbitrarily modify news headlines *after* their initial release. The motivation, frequency, and effect of post-publication headline changes are largely unknown, with no offline equivalent from where researchers can draw parallels.

In this paper, we collect and analyze over 41K pairs of altered news headlines by tracking ~411K articles from major US news agencies over a six month period (March to September 2021), identifying that 7.5% articles have at least one post-publication headline edit with a wide range of types, from minor updates, to complete rewrites. We characterize the frequency with which headlines are modified and whether certain outlets are more likely to be engaging in post-publication headline changes than others. We discover that 49.7% of changes go beyond minor spelling or grammar corrections, with 23.13% of those resulting in drastically disparate information conveyed to readers. Finally, to better understand the interaction between post-publication headline edits and social media, we conduct a temporal analysis of news popularity on Twitter. We find that an effective headline post-publication edit should occur within the first ten hours after the initial release to ensure that the previous, potentially misleading, information does not fully propagate over the social network.

CCS CONCEPTS

• **Information systems → Association rules**; • **Security and privacy → Social aspects of security and privacy**.

KEYWORDS

Information Integrity, News Title Modification Taxonomy, Information Propagation over Social Network

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '22, April 25–29, 2022, Virtual Event, Lyon, France.
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9096-5/22/04...$15.00
https://doi.org/10.1145/3485447.3512219

ACM Reference Format:
Xingzhi Guo, Brian Kondracki, Nick Nikiforakis, and Steven Skiena. 2022. Verba Volant, Scripta Volant: Understanding Post-publication Title Changes in News Outlets. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3485447.3512219

1 INTRODUCTION

The growth of the Internet has shifted the distribution of news articles from traditional channels such as print and radio to online websites and social media platforms. Today, 86% of people consume news on digital devices via the Internet [29]. This new medium has greatly reduced the time between the occurrence of newsworthy events and the publishing of articles reporting on them. One of the strategies that news agencies appear to be regularly using is to publish articles with incomplete information, taking advantage of their ability to modify online articles *after* their publication. Consequently, readers who view an article or headline immediately after publication may be exposed to radically different information and therefore arrive at different conclusions, compared to readers consuming the same article at a later time. For example, *The Atlantic* published the following article on February 11th, 2021 with the title:

Original title[21]: *"I Miss the Thrill of Trump"*

However, five hours later the title of the article changed to:

Altered title[22] : *"I Was an Enemy of the People"*

Given the documented tendency of users to skim through titles rather than read the content of every news article that they encounter [6], a radical modification of a headline translates to two or more groups of people who have consumed the same news from the same sources, yet arrive at potentially different worldviews.

In this paper, we explore the phenomenon of post-publication modifications of news article headlines. We monitor the titles of articles from several top news publishers, and capture all post-publication changes for later analysis. We analyze our compiled dataset to create a taxonomy of title changes using an automated NLP pipeline allowing us to determine the reasons for article title changes, the entities responsible, and the tangible benefits the publisher may receive for doing so. Further, we examine the spread of news over social networks by collecting time-series impression statistics (e.g., number of retweets/likes) of tweets containing links to news articles for the Twitter accounts of news publishers.

The primary contributions of our work are as follows:

- *Temporal News Headline Data Set*: We have developed a temporal news headline corpus consisting of 30,930 distinct news articles where the headline changed at least once. Our dataset of 41,906 total altered headlines pairs (some articles change headlines more than once) provides a unique resource for future studies of news trustworthiness as well as user trust, which we will make available upon publication.
- *Characterizing Changes in News Headlines*: We create an NLP categorization pipeline for assessing headline changes by the proposed nine-class taxonomy, based on journalism domain knowledge and a state-of-the-art language model (BERTScore [34]), from which we successfully discover that 23.13% headline edits are perceived as harmful. This analysis enables us to quantify why headline changes occur in practice, and how policies differ with particular news agencies.
- *Estimating Effective Timing over Twitter*: We present a temporal analysis showing how rapidly news is fully propagated over social networks. We observe that most news tweets are shared/retweeted within the first 10 hours following their initial posting, before gradually fading out due to losing the public's attention, or achieving their maximum audience. This suggests that an effective headline correction must occur quickly to avoid propagating misinformation.

2 RELATED WORK

A genuine news headline summarizes the content and enables readers to draw quick conclusions [10, 15, 26]. However, in the digital media era, false information or fake news [11, 19, 30, 35] threatens the public's information consumption by inducing readers with clickbait headlines and fabricated content. Vargas et al. [31] developed techniques to distinguish legitimate activity on Twitter from disinformation campaigns using coordination network analysis. Additionally, a large research line [12, 16, 25, 35] focus on fake news detection, satire detection[9, 28], and clickbait detection [23]. Hounsel et al. [14] proposed methods to discover disinformation websites using characteristics of their hosting infrastructure. In this work, we investigate news outlets that, in the process of publishing legitimate articles, modify their headlines.

The headline modifications may be harmless (e.g., updating competition score) or malicious (e.g., making the headline clickbaity [7], or starting with an inaccurate headline that maximizes user views and eventually changing it to an accurate one). Hagar and Diakopoulos [13] discussed A/B testing on news headlines and gather audiences' feedback on several headline writing practices (e.g., starting headlines with "why" or "how" and subjective ideals) and reported that headlines are optimized for specific metric (e.g., click-through rate) from A/B testing. Kuiken et al. [18] investigated the effectiveness of headlines by comparing click-through rate of the original title and its of the rewritten one, suggesting that clickbait features led to statistically significant increase in number of clicks. However, the size of their dataset was limited, with 1,836 pairs of headlines that were rewritten from a single groups of editors.

Another related research line is text edit classification. Previous work [8] analyzed linguistic features for the task of edit category classification, organizing English Wikipedia's edits into 21 categories,

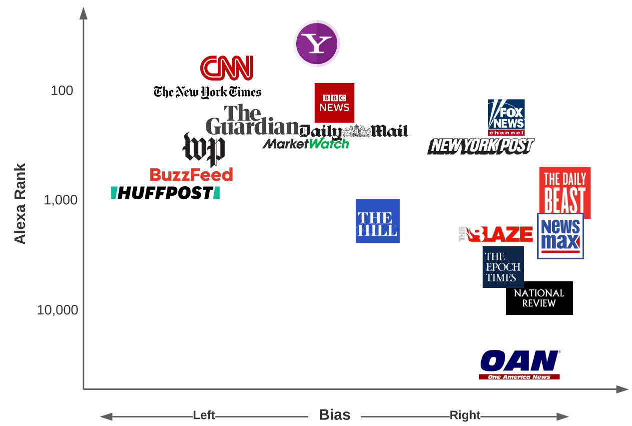

Figure 1: News publishers chosen based on their political bias and Alexa ranking.

from spelling corrections to vandalism. Yang et al. [32] investigated the intentions behind text edits and created a 13-category taxonomy of edit intentions, then developed supervised learning models for automatic identification. Similarly, Marrese-Taylor et al. [20], Yin et al. [33] employed encoder-decoder deep learning framework to learn the edit representation and predict its categories. Due to data availability, previous research mainly focus on Wikipedia post-edit instead of news headlines. Given our unique, large-scale dataset of 41K pairs of title changes, we aim to understand real-world headline modifications and derive a taxonomy for automated edit categorizations. In addition, we analyze the speed of news propagation over social networks, how they relate to post-publication headline changes, and offer guidance on the timings of responsible post-publication modification.

3 DATASET PREPARATION

In this section, we discuss our process for selecting news publishers to include in our study. We then describe the infrastructure we designed and implemented to capture news articles, as well as detect post-publication headline changes and measure the spread of news on social media.

3.1 News Publisher Identification

Prior to capturing and studying post-publication article headline changes, we must first identify a set of news publishers to collect article data from that are representative of a broad range of audiences, biases, and platform sizes. To this end, we constructed a set of news publishers by first consulting the Media Bias Chart created by Ad Fontes Media [2]. This chart maps news publishers onto a two-dimensional plane representing political bias, and reputability. We focus only on the political bias of each news publisher, selecting an equal number of publishers from each region of this axis.

Rather than utilize the reputability ranking of each publisher on this chart, we opt instead to measure reputability with the Alexa ranking [1] of each publisher's website. We reason that the Alexa ranking of each publisher is an effective and unbiased proxy for a publisher; with more reputable publishers drawing a larger audience

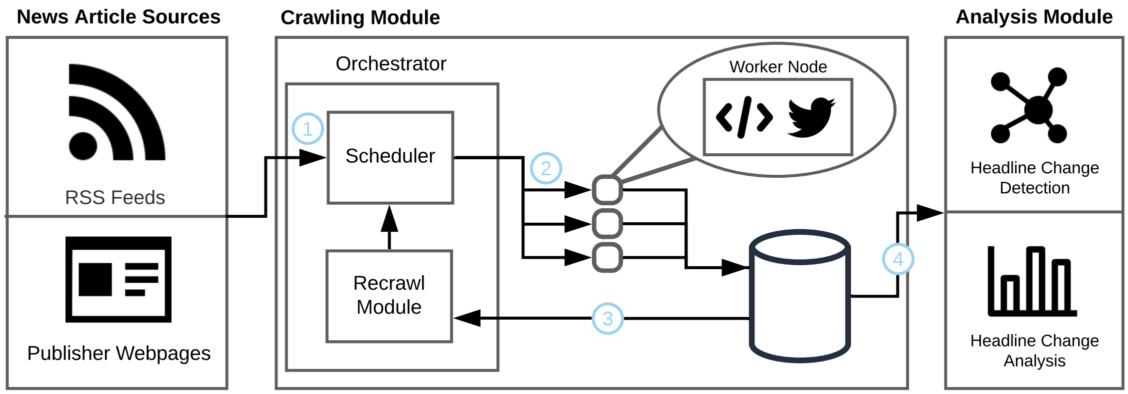

Figure 2: The overview of news crawling infrastructure

and higher rank. We group publishers into buckets representing different Alexa rank ranges aiming for an equal number of publishers from each bucket. Figure 1 shows the news publishers we chose, and their positions in the two-dimensional plane we described.

3.2 Data Collection Infrastructure

To measure the extent in which news publishers modify the titles of their articles post-publication, we created a data collection infrastructure that visits the web pages of news articles as they are posted and continually monitors them to detect changes to their titles and content. Additionally this infrastructure monitors the spread of news on Twitter.

Post-publication Article Headline Changes Figure 2 provides an overview of our infrastructure. (i) News article URLs are collected using a set of input modules corresponding to the source in which the URLs are gathered from. For all news publishers we study, URLs are collected using the RSS feeds provided by each publisher. However, our infrastructure can be easily extended to support additional URL sources, such as scraping the home page of a news publisher, by creating a new module for that source. Our RSS module queries the RSS servers of each news publisher periodically to gather the URLs of new articles shortly after they are posted.

The URLs for each new article are placed onto a queue in the article crawling module (ii) where workers consume new URLs by visiting each URL and parsing metadata from the article web page. We collect and save each HTML web page for future processing, but specifically parse out the article title using the Python Newspaper library [4].

In order to detect title changes in news articles, (iii) we recrawl each article web page periodically for the two days following its original publication and parse out the same information as previously described. We arrived at this two-day threshold via a pilot crawling experiment where we established that any headline modifications were typically occurring in the first few hours after an article was published.

Measuring the Spread of News on Twitter In many cases, modifications to news articles occur *after* these articles have already been consumed and shared. Social media platforms only exacerbate this problem due to the speed in which information propagates among users. We chose to study the propagation of news on Twitter because, unlike other social media platforms, Twitter prevents users from editing tweets after they are posted, leading to discrepancies in the information shared in the tweets of news agencies, and their related articles. When crawling an article found to have a modified title, our infrastructure also searches the Twitter feed of the relevant news publisher for a tweet containing a link to the article in question. If such a tweet is discovered, it is recrawled at the same interval as the article itself. For each crawl of a tweet, our infrastructure records the tweet text, as well as the current number of favorites and retweets. To measure the base level of interaction tweets from each news publisher receive, our data collection infrastructure also crawls a random sample of tweets from each publisher.

4 POST-PUBLICATION HEADLINE CHANGES

Using our data collection infrastructure, we monitored 411,070 articles published by the news agencies listed in Figure 1 from March 1st, 2021 to August 31st, 2021. In total, we observed 30,930 (7.5%) articles modify their headlines at least once, resulting in 41,906 changed pairs[1]. Figure 3 shows that over 90% of changes occur within the first 10 hours after publication. Readers who encounter an article shortly after publication will likely come away with a different opinion on the subject-matter than a reader who encounters the same article after the headline has changed. Furthermore, if the headline change is to correct invalid information, this can lead to rapid spread of misinformation through channels such as social media. In this section, we explore the magnitude of headline changes in our dataset, assign labels to these changes corresponding to the nature of the modification, and compare the consistency of headlines published by popular news outlets.

4.1 Headline Change Magnitude

The inclusion, exclusion, or modification of a single word or phrase in a headline can lead to large differences in reader perception. Figure 3 shows the edit distance of headline modifications we detected during our data collection period. In the context of headline modifications, we define a token as a single word. We observe a bi-modal

[1]We consider one pair as two consecutive headline versions. Additionally, we list news categories of changed pairs per agency in Table 8 in Appendix

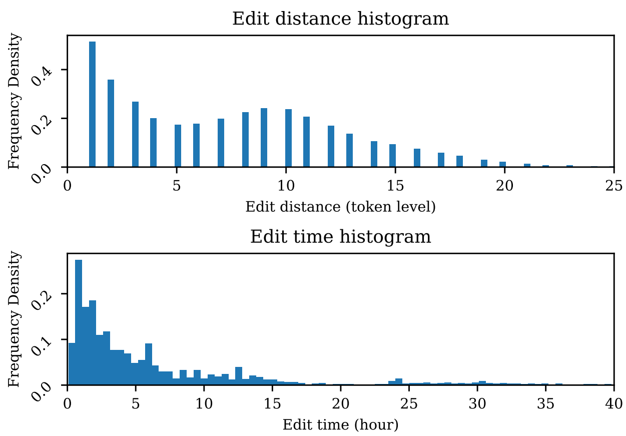

Figure 3: Edit distance histogram (*top*): a large fraction of edits are local edits with distance 1 to 3. Timing of title changes (*bottom*): Most title changes occur within 10 hours after initial publishing. We present plots per publisher in Figure 8 in the appendix.

distribution in our dataset, with a majority of headline changes resulting in an edit distance of a single token, and a second local peak at approximately 10 tokens. Since we expect headline modifications in these two groups to be the result of distinct contributing factors, we discuss them separately.

Single-token Edit Distance. During our data-collection period, we observed 6,036 single-token edit changes. Among them, 3,767 are word substitution. Table 1 shows the top five most common part-of-speech (POS) changes in headlines . We find that headlines with single-token substituted typically correspond to either updates of ongoing events, or error corrections.

Table 1: The top-5 changes in part of speech

POS Changes	#
(NOUN, NOUN)	861
(VERB, VERB)	573
(NUM, NUM)	396
(PUNCT, PUNCT)	242
(ADJ, ADJ)	216
Others	1,479

Specifically for word substitution, 44.1% of noun changes and 54.6% of verb changes are synonyms, hyponymns, hypernymns, or otherwise share the same lemma. Tables 6 and 7 in the appendix list these breakdowns for nouns and verbs, respectively. These modifications typically occur at the end of a newsworthy event; with modified verbs transitioning from present to past tense, indicating the end of the event. Similarly, changes in headlines correspond to updates in ongoing events (e.g. score changes in sporting events). We observe that 68.39% of numerical headline changes result in the number increasing by some value (e.g. reflecting more discovered injuries/casualties from an ongoing natural disaster). Table 5 in the appendix shows the full breakdown of all numerical changes we observed.

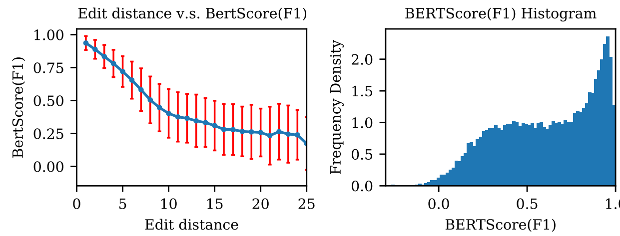

Figure 4: Figure *(left)*: BERTScore decreases as more edits being applied until the pre/post-edit titles become irrelevant (BERTScore reaches to 0). Figure *(right)* shows that there is only few complete rewrite in news titles.

Arbitrary Edit Distance. To model headline modifications of arbitrary edit distance, we compute the difference of lexicon in changed pairs to distinguish whether it is a minor update or a complete rewrite. BERT-based word alignment score –*BERTScore* [34] as defined in Eq.1 – is a lexical alignment statistic using contextual word embeddings with attention mechanism. BERTScore can capture the semantic similarity in two sentences of dissimilar structures and lexicons, whereas an edit distance only measures superficial lexical differences. We extend our analysis to all changed pairs with $BERTScore_{F1} \in [-1,1]$ as the semantic similarity measure. The relationship between edit distance and $BERTScore_{F1}$ is shown in Figure 4 *(left)*.

$$R = \frac{1}{|x|}\sum_{x_i \in x} max_{\hat{x}_j \in \hat{x}} x_i \hat{x}_j;\ P = \frac{1}{|\hat{x}|}\sum_{\hat{x}_j \in \hat{x}} max_{x_i \in x} x_i \hat{x}_j$$
$$BERTScore_{F1} = 2\frac{P * R}{R + P} \in [-1,1], \tag{1}$$

where x_i,$\hat{x}_i$ denote i^{th} token in two sentences x,$\hat{x}$.

Figure 4 *(right)* shows the distribution of BERTScore where the peak around 0.9 indicates that most edits do not change the semantic meaning of the headline, while the samples centered around 0 represent the post-publication edit completely changing the headline's semantic meaning. We present examples associated with BERTScore in Table 9 of the appendix. Using a BERTScore threshold of 0.25, we find approximately 10% of the changed pairs in our dataset are significantly rewritten. We consider these headline changes to be the most damaging to readers as it is very likely that the message conveyed by the headline will change drastically after modification.

Table 2 shows the ordered list of news agencies by the average similarity score across all changed headlines. Generally, higher similarity indicates minor changes that do not alter the original semantic meaning. We find many well-regarded publishers (e.g., BBC, The Guardian, NYT) frequently modify headline semantics. This unexpected finding suggests that the popularity of a news outlet does not necessarily indicate greater restraint with post-publication headline changes. Contrastingly, we observe publishers such as The Epoch Times with an Alexa rank of close to ten thousand modify headlines at a much lower rate than publishers with a much larger online presence.

Additionally, by noting the ratio of modified headlines with the average BERTScore for all modified headlines, we can deduce the

Table 2: The statistics of semantic similarity (BERTScore) by agency (Ranked by the mean value). To clarify the similarity rankings from news reputation rankings, the ordered list does not reflect credibility. *Tracked* is the total number of tracked articles, *Mod. Ratio* is the percentage of the articles with at least one altered headline.

Agency	Mean	Median	Tracked	Mod.Ratio
Huffington Post	0.8292	0.8966	4,875	0.0357
The Epoch Times	0.8260	0.8908	4,765	0.1442
The Hill	0.8087	0.8644	5,676	0.0592
Fox News	0.7659	0.8645	11,003	0.0486
New York Post	0.7305	0.8174	13,881	0.0665
National Review	0.7268	0.8305	3,324	0.0126
The Blaze	0.7200	0.8145	4,401	0.0400
CNN	0.6888	0.7493	6,344	0.2030
Washington Post	0.6840	0.7916	6,076	0.1508
Newsmax	0.6733	0.7319	1,115	0.0233
Daily Beast	0.6584	0.7693	4,359	0.1308
MarketWatch	0.6551	0.7011	7,503	0.3453
BBC	0.6031	0.6260	5,854	0.2277
Yahoo News	0.5985	0.6160	227,246	0.0846
Daily Mail	0.5936	0.6348	75,118	0.0756
OAN	0.5830	0.5910	8,323	0.2616
The Guardian	0.5579	0.5826	6,916	0.1975
New York Times	0.5234	0.5595	9,746	0.1892
BuzzFeed	0.5085	0.4872	4,545	0.4438

expected consistency of a particular news outlet. For instance, we find that BuzzFeed modified the headlines of 44% of their articles during our data-collection period. Of those articles, we observe the headline changes produce the lowest average BERTScore (i.e. the modified headlines depart significantly from the original ones). Contrastingly, during our data collection period, The Huffington Post only modified the headlines of 3% of observed articles; achieving the highest average BERTScore with those changes. We can therefore conclude that, in terms of headline modifications, The Huffington Post is a more consistent news outlet compared to BuzzFeed. We note however, that this consistency of headlines does not necessarily imply accuracy of headlines.

4.2 Categorization of News Title Changes

To discover trends in article headline changes, and the behaviors of the publishers responsible, we create a nine-class taxonomy of editing types based on journalism knowledge (e.g. Emotionalism, etc.), existing taxonomies for Wikipedia edits[8, 32], and the most common edits from our observations (e.g., Paraphrase, Dynamic update, etc.) . We design our taxonomy as follows:

- **Paraphrase:** We assign the label *Paraphrase* if the changed pairs have a similarity score (BERTScore-F1) greater than 0.8.
- **Dynamic Update:** If both old and new title contain keywords such as *Monday brief, update, live, stream, etc.* The keywords were manually selected.
- **Emotionalism[17]:** if only the news title contains any of words in the subjective dictionary.

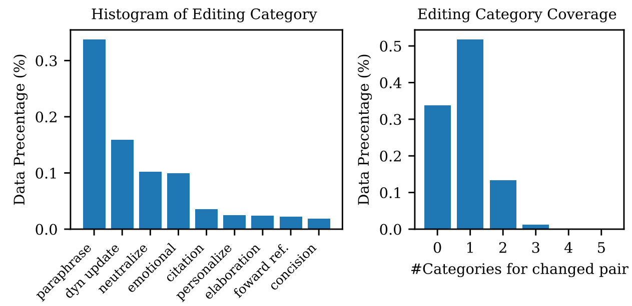

Figure 5: Assigned category statistics. Figure (*left*) shows the majority change types are paraphrase and dynamic update. Figure (*right*) shows that approx. 70% of samples have at least one category assigned. 50.3% modifications associate with benign edit categories (paraphrase, updates, elaboration, concision), while 23.13% modifications are linked with the other less benign edits.

- **Neutralization:** if only the old title contains subjective words mentioned above.
- **Forward Reference:** Forward Reference[7, 17] is a common feature in news titles, piquing reader curiosity. We assign this label only if the changed title contains the keywords such as *why, when, which, how, etc.*
- **Personalization[17]:** Personalization is used to retain audiences and make the readers feel involved in the news. We assign this label only if the new title contains keywords such as *you, we, s/he, your, etc.*
- **Citation:** Citation is a common technique in news headlines that make them look more reliable. We assign this label only if the new title contains keywords like *said, says, told, etc.*
- **Concision:** If the new title removes some text, and the remaining text aligns with the old title ($P < 0.6$, $R > 0.5$). BERT-Score Precision/Recall is defined in Eq.1
- **Elaboration:** If the new title adds text, and is semantically aligned with the old title: $P > 0.6$, $R < 0.5$).

These nine categories allow us to better understand headline modifications by classifying them into groups corresponding to their perceived purpose. With the combination of BERTScore, Stanza Pipeline [24], word sentiment dictionary [27], and hand-crafted rules, we are able to automatically assign labels using our proposed NLP pipeline. We empirically determined the thresholds to use by finding those that made the most sense with manual inspection. Table 9 of the appendix present more examples of each of these categories.

Figure 5 *(right)* shows the resulting label coverage. Our proposed categorization rules cover approximately 70% of news title changes (the changed pair has at least one assigned category). Among the 30% of modifications not covered by our pipeline, we find that the missing cases usually have a BERTScore of less than 0.8, involve more complex sentence structure changes (double negation), or out-of-vocabulary words (most of them are COVID-related "COVID-19, J.&J., AstraZenec", which debuted after the release of the pre-trained model). We listed several examples in Table 3 and more examples in Table 9 in the Appendix.

Table 3: Examples of the altered titles sampled from our dataset by the modification category. The example of the label "Other" involves out-of-vocabulary words (J.&J) and changes in tone. The example of "Concision" removes the specific person, while the "Elaboration" example expands the original title with an attributive clause. "Forward-Reference" adds a question mark and "Personalization" substitute reader into the title, which both trigger curiosity from the readers. The "Neutralize" example remove *"long"* from the original title, making it less subjectivity. While "Emotional" example adds *"tease, big"*, which makes the title more inflammatory.

Before	After	BERT-F1	Label
One Dose of J.&J. Vaccine Is Ineffective Against Delta...	J.&J. Vaccine May Be Less Effective Against Delta ...	0.6830	Other
Senator Capito says Republicans plan new US infrastructure offer	U.S. Senate Republicans prepare new infrastructure offer	0.5434	Concision
Raul Castro confirms he's resigning, ending long era in Cuba	Raul Castro resigns as Communist chief, ending era in Cuba	0.7310	Neutralize
Fourth stimulus check update: Your next payment could be these	Fourth stimulus check? These payments are already in the pipeline	0.4479	Forward Ref.
The Latest: UN: 38,000 Palestinians displaced in Gaza	The Latest: Biden expresses 'support' for Gaza cease-fire	0.4043	Dyn. Update
MAGA 2.0	"A new bargain": Biden's 2024 tease bets big on nostalgia	0.0257	Emotional
23 Amazing Jokes From Hot Fuzz	23 Jokes From "Hot Fuzz" That Humans Will Laugh At For The Next 10,000 Years	0.4715	Elaboration
Watch Jeff Bezos' 'Blue Origin' launch into space live today	Everything you need to know as Jeff Bezos' 'Blue Origin' launches into space today	0.6179	Personalize
Afghan guard killed: Firefight leaves at least one dead and others injured at Kabul airport	'It would be better to die under Taliban rule than face airport crush', say US embassy's 'betrayed' Afghan staff	0.0656	Citation

Table 4: The breakdown of edit types per news agency. Each headline modification may belong to multiple labels. We highlight the top one in each column.

Agency	Paraphrase	Dynamic Update	Elaboration	Concision	Emotional	Neutralize	Forward Reference	Personalize	Citation
BBC	0.3196	0.3848	0.0278	0.0113	0.1043	0.0818	0.0180	0.0188	0.0540
BuzzFeed	0.1587	0.0203	0.0312	**0.1101**	0.1026	**0.1894**	**0.0709**	**0.0699**	0.0238
CNN	0.4262	0.0916	0.0404	0.0148	0.0699	0.0738	0.0171	0.0342	0.0435
Daily Beast	0.4754	0.1035	0.0246	0.0175	0.1123	0.0912	0.0386	0.0351	0.0298
Daily Mail	0.2801	0.1158	0.0357	0.0144	0.0993	0.1218	0.0241	0.0521	0.0422
Fox News	0.6056	0.1925	0.0187	0.0093	0.0710	0.0617	0.0168	0.0224	0.0262
Huffington Post	**0.7356**	0.0690	0.0000	0.0000	0.0345	0.0632	0.0000	0.0345	0.0230
MarketWatch	0.3867	0.0872	0.0413	0.0178	0.0965	0.0799	0.0363	0.0154	0.0247
National Review	0.5476	0.3810	0.0476	0.0000	0.0714	0.0476	0.0000	0.0238	0.0000
New York Post	0.5211	0.0997	0.0336	0.0076	0.0715	0.0618	0.0336	0.0293	0.0141
New York Times	0.2923	0.1171	0.0108	0.0108	0.1123	0.1432	0.0380	0.0184	0.0331
Newsmax	0.3846	0.3077	0.0385	0.0000	0.0000	0.0769	0.0000	0.0000	0.0385
OAN	0.2692	0.0156	0.0280	0.0184	0.1029	0.0988	0.0032	0.0023	0.0271
The Blaze	0.5341	0.3239	**0.0625**	0.0227	0.0682	0.0398	0.0227	0.0398	0.0455
The Epoch Times	0.6856	0.1659	0.0146	0.0204	0.0277	0.0422	0.0044	0.0087	0.0364
The Guardian	0.2950	**0.4100**	0.0271	0.0124	**0.1215**	0.1164	0.0190	0.0190	**0.0688**
The Hill	0.6637	0.2024	0.0268	0.0327	0.0208	0.0446	0.0089	0.0060	0.0179
Washington Post	0.4891	0.2205	0.0186	0.0109	0.0884	0.0797	0.0207	0.0437	0.0371
Yahoo News	0.3256	0.1846	0.0150	0.0126	0.1047	0.0977	0.0156	0.0161	0.0340

We also find that the most common changes belong to the "Paraphrase" and "Dynamic Update" categories (44.24%), which do not significantly change the semantic meaning of the headlines. Changes in these categories typically correspond to updates of ongoing events and grammatical/spelling corrections. However, we note that all other categories in our taxonomy may contain headline changes that can be perceived as harmful to at least some readers.

Over the course of our data collection period, we observed an equal distribution of "Emotionalism" and "Neutralization" headline changes, suggesting that publishers are just as likely to add emotional words to attract more readers after initial publication as they are to remove them. However, since both of these categories transition the headline from a provocative to non-provocative state, or vice-versa, these two groups can have the same effect on different groups of readers. In the case of "Emotionalism" changes, readers who view a headline after it has been modified, will likely have a more emotional response to the subject-matter, while "Neutralization" changes will have the same effect on readers who view a headline pre-modification. Together, these two groups make up 20.13% of all headline changes in our dataset.

Table 4 shows the percentage of articles by each publisher belonging to each of the nine categories in our taxonomy. Since it reflects the explanation of headline changes, the percentage does not reflect the general media bias, which may be embedded in the static news headlines. We observe that over 70% of the articles published by the Huffington Post, The Epoch Times and The Hill belong to either the "Paraphrase" or "Dynamic Update" categories, demonstrating that their headlines will typically not change in ways which will lead to drastically different opinions among readers. We also notice that *BuzzFeed* leads in three change categories, including the categories most closely related to click-bait: "Forward Reference", and "Personalization".

By observing the categorizations of each publisher, a model can be developed on their overall headline-modification strategies. For instance, the "Citation" category of our taxonomy allows us to deduce that *The Guardian* is the most likely to include witness or expert testimony in headlines post-publication, with 6.8% of their articles falling into that category. Conversely, we find that the publisher National Review is very unlikely to make such a headline change as we did not observe any instances of modification by either publisher during our data collection period. These observations quantify the regularity with which publishers modify their headlines, and can be used as a supplementary feature for future research on digital journalism and public trust in news.

4.3 News Propagation over Twitter

In recent years, social media services have emerged as a vital tool for news publishers to quickly share stories with the public. Due to the ability of users to share and "repost" links to articles with others connected to them online, information can spread faster than ever before. Because of this user-assisted amplification, false or misleading information in article headlines can propagate to exponentially many readers before headline changes occur, magnifying the negative repercussions of unanticipated headline changes.

We track the reachability of news on Twitter in terms of user engagement. As mentioned in Section 3, our data-collection infrastructure records the number of likes and retweets of tweets associated with news articles with modified headlines from the accounts of each publisher. In total, we detect 5,384 tweets corresponding to articles with modified headlines during our data collection period. We note that these tweets correspond to the original article headlines, with their immutability preventing publishers from simply editing their contents to reflect headline changes. Rather, publishers must first delete the original, outdated tweet, and replace it with a new tweet. Alarmingly however, we find that only 0.15% of all such tweets were deleted subsequent to a headline change in the corresponding article. Moreover, we find that these tweets on average garner 39 retweets and 137 favorites. This means that the vast majority of tweets associated with outdated and potentially misleading information remain

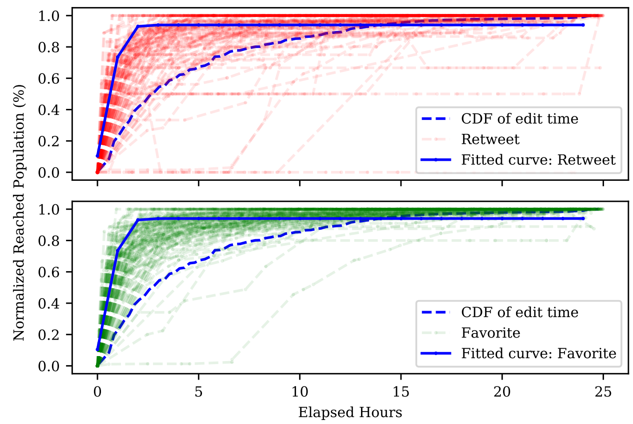

Figure 6: The normalized retweets (top) and favorites (bottom) V.S. news published time. The solid blue curve is the fitted curve $y = a \cdot (1 + e^{-b(x-c)})^{-1}$, approximating how quick the news achieves its max influence.

online indefinitely, and receive considerable attention from users who then spread this information to others.

Figure 6 shows the distribution of time in which tweets associated with headline changes were "retweeted" and "favorited" by users. We observe a clear trend that news stories gradually lose popularity on social networks as they become less novel. People usually share (retweet) tweets associated with news stories within the first ten hours and react (favorite) within first five hours. Based on the observations in Figure 3, we found that almost 80% title changes happened within the first five hours after publishing. These findings are alarming as they show that many headline changes occur *after* most social media activity regarding the particular article ceases. This means that social media users are currently either spreading false or misleading information before the publisher can remedy it by modifying the headline, or sharing articles that will portray an entirely different emotional response when their followers view it at a later point in time. Both of these scenarios can lead to situations where different readers of the same article or headline will come to different conclusions depending on when they encountered it.

URL Preview Cards A popular feature of Twitter is the rendering of cards that preview the content located at the URL linked in a tweet. These cards work by parsing and rendering the Open Graph [3] metadata from the HTML source of the linked webpage directly below the tweet text. Typically, a webpage thumbnail, title, and description are rendered. This metadata is cached for each particular URL for approximately one week; only updating when the URL is modified [5]. This caching behavior can amplify the negative repercussions of post-publication headline changes as any tweet containing the URL of an article that was published prior to a headline change will display the outdated headline for at least one week. This is a behavior we have observed on the Twitter accounts of popular news publishers. Figure 7 demonstrates an example of this behavior with a headline change from the BBC, along with a tweet containing a link to the article posted by the official BBC Twitter account. Twitter users who only read the headline and description located on the rendered card

Figure 7: Twitter URL preview of article posted by BBC Politics. Article headline outdated on Twitter due to caching behavior of URL cards.

will consume outdated information on the situation, as opposed to those who click through to the article webpage.

5 DISCUSSION

Headline Changes. The Internet has dramatically decreased the time between newsworthy events, and the consumption of information regarding those events by the public. This, along with the propensity of people to simply skim headlines rather than read full articles [6], has fostered an environment where post-publication changes to article headlines can lead to diverging world-views among readers. In this work, we studied the behavior of news publishers and developed a taxonomy of article headline modifications. We find that popular news publishers regularly make post-publication changes to their headlines, with some modifying almost half of them.

Additionally, we categorize these changes into well-defined groups, allowing us to quantify the behaviors of each publisher. We find that the majority of headline changes correspond to updates of ongoing events. Alarmingly however, 20.13% of headline changes are made to either increase or decrease the provocativeness of the headline, presenting a distorted view of the subject-matter to different groups of readers. Using this taxonomy, we are able to quantify the different headline update strategies of each news outlet studied, discovering a divergence in the perceived motives between publishers.

While the news outlets presented in this study do not appear to have outright malicious intentions, their actions contribute to the overall decrease in integrity of information consumed by millions of people. In order to slow the spread of misinformation in society, news outlets should limit the rate in which their article headlines are modified. With the speed in which news travels, headline changes can have devastating effects to reader understanding. We argue that, if a headline must be changed, the new headline should maintain high lexical similarity to the original headline, only adding new information to enhance reader understanding.

Information Travel on Social Media. The rapid propagation of information through shares and retweets on social media has only exasperated the negative ramifications of news article headline modifications. By observing the engagement of tweets associated with news articles with modified headlines, we determined that the majority of favorites and retweets occur within ten hours of a tweet's publication. Due to the caching of article metadata and previews on Twitter, most article shares will contain a different headline than what is active on the publisher's website. While it is infeasible for social media platforms to constantly monitor each and every webpage linked to from all posts, decreasing the caching of content can help reduce the spread of misinformation online. Moreover, it would be worthwhile to explore ideas around the penalization of aggressive headline changes. That is, if an article received 10K retweets *before* a significant headline change, is it appropriate for that article to keep all the clout that the previous title generated? Penalizing large changes has the potential to act as a deterrent of unwanted publishing behavior and encourage publishers to be more considerate about their content choices *before* an article is published.

6 CONCLUSION

In 2022, trust in the media is at an all-time low. In this paper, we explored one dimension that we argue has the potential to further reduce the public's trust in news outlets: post-publication headline changes. By monitoring over 411K articles for seven months across tens of news outlets, we discovered that 7.5% of titles changed at least once after they were published. This rate of headline modifications is anything but uniform, with certain popular outlets changing almost half of their headlines *after* publication. We used the BERTScore metric and devised a taxonomy to automatically characterize the type of post-publication change. We find that 49.7% of changes go beyond benign corrections and updates, with 23.13% corresponding to categories such as emotionalism, neutralization, and personalization. We also characterized the effects of post-publication headline changes in relation to social media. Among others, we discovered that maximal spread of news on Twitter happens within ten hours after publication and therefore a delayed headline correction will come *after* users have consumed and amplified inaccurate headlines. Finally, we discussed the issue of content caching in social networks and how it can further exacerbate the propagation of stale headlines.
Acknowledgements: This work was supported by the National Science Foundation (NSF) under grants CMMI-1842020, CNS-1941617, CNS-2126654, IIS-1926781, IIS-1927227, IIS-1546113 and OAC-1919752.
Availability To assist in the understanding of the spread of misinformation on the web, we are open-sourcing our news headline dataset: `https://scripta-volant.github.io/`

REFERENCES

[1] [n. d.]. Alexa Top Sites. https://www.alexa.com/topsites.
[2] [n. d.]. Media Bias Chart. https://www.adfontesmedia.com/intro-to-the-media-bias-chart/.
[3] [n. d.]. Open Graph Protocol. https://ogp.me.
[4] [n. d.]. Python Newspaper Library. https://newspaper.readthedocs.io/en/latest/.
[5] [n. d.]. Twitter Card Troubleshooting. https://developer.twitter.com/en/docs/twitter-for-websites/cards/guides/troubleshooting-cards#outdated.
[6] 2014. How Americans get their news. https://www.americanpressinstitute.org/publications/reports/survey-research/how-americans-get-news/.
[7] Jonas Nygaard Blom and Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76 (2015), 87–100.

[8] Johannes Daxenberger and Iryna Gurevych. 2013. Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 578–589.

[9] Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. Attending sentences to detect satirical fake news. In *Proceedings of the 27th international conference on computational linguistics*. 3371–3380.

[10] Daniel Dor. 2003. On newspaper headlines as relevance optimizers. *Journal of pragmatics* 35, 5 (2003), 695–721.

[11] Lisa Fazio. 2020. Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review* 1, 2 (2020).

[12] Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jennine B Everett, et al. 2018. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science*. 17–21.

[13] Nick Hagar and Nicholas Diakopoulos. 2019. Optimizing content with A/B headline testing: Changing newsroom practices. *Media and Communication* 7, 1 (2019), 117–127.

[14] Austin Hounsel, Jordan Holland, Ben Kaiser, Kevin Borgolte, Nick Feamster, and Jonathan Mayer. 2020. Identifying disinformation websites using infrastructure features. In *10th {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 20)*.

[15] Elly Ifantidou. 2009. Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics* 41, 4 (2009), 699–720.

[16] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*. 1546–1557.

[17] Danielle K Kilgo and Vinicio Sinta. 2016. Six things you didn't know about headline writing: Sensationalistic form in viral news content from traditional and digitally native news organizations. In *Research Journal of the International Symposium on Online Journalism*, Vol. 6. 111–130.

[18] Jeffrey Kuiken, Anne Schuth, Martijn Spitters, and Maarten Marx. 2017. Effective headlines of newspaper articles in a digital environment. *Digital Journalism* 5, 10 (2017), 1300–1314.

[19] Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559* (2018).

[20] Edison Marrese-Taylor, Machel Reid, and Yutaka Matsuo. 2020. Variational Inference for Learning Representations of Natural Language Edits. *arXiv preprint arXiv:2004.09143* (2020).

[21] Alexander Nazaryan. 2021. I Miss the Thrill of Trump. https://web.archive.org/web/20210211110734if_/https://www.theatlantic.com/ideas/archive/2021/02/i-miss-thrill-trump/617993/.

[22] Alexander Nazaryan. 2021. I Was an Enemy of the People. https://web.archive.org/web/20210223004426if_/https://www.theatlantic.com/ideas/archive/2021/02/i-miss-thrill-trump/617993/.

[23] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *European Conference on Information Retrieval*. Springer, 810–817.

[24] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082* (2020).

[25] Julio CS Reis, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. 2019. Supervised learning for fake news detection. *IEEE Intelligent Systems* 34, 2 (2019), 76–81.

[26] Julio Rieis, Fabrício de Souza, Pedro Vaz de Melo, Raquel Prates, Haewoon Kwak, and Jisun An. 2015. Breaking the news: First impressions matter on online news. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 9.

[27] Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*. 105–112.

[28] Victoria L Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection*. 7–17.

[29] Elisa Shearer. 2021. More than eight-in-ten Americans get news from digital devices. https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/.

[30] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[31] Luis Vargas, Patrick Emami, and Patrick Traynor. 2020. On the detection of disinformation campaign activity with network analysis. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*. 133–146.

[32] Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy. 2017. Identifying semantic edit intentions from revisions in wikipedia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2000–2010.

[33] Pengcheng Yin, Graham Neubig, Miltiadis Allamanis, Marc Brockschmidt, and Alexander L Gaunt. 2018. Learning to represent edits. *arXiv preprint arXiv:1810.13337* (2018).

[34] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675* (2019).

[35] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 836–837.

A APPENDIX

Table 5: Numerical difference between pre/post-edits in changed news titles: Most of them concern news dynamic updates of injuries and competition scores. High *BERTScore* shows that numerical changes does not alter the meaning.

Diff.	Number	Before	After	BERTScore
1.0	116	louisiana floods lead to 6 deaths	louisiana floods lead to 7 deaths	0.99
2.0	45	yankees lead astros 1-0 in game 1: live score and updates	yankees lead astros 3-0 in game 1: live score and updates	0.96
-1.0	31	ransomware attack hits 23 texas towns, authorities say	ransomware attack hits 22 texas towns, authorities say	0.99
3.0	24	super bowl 2020 live: chiefs lead the 49ers, 7-3	super bowl 2020 live: chiefs lead the 49ers, 10-3	0.93
4.0	17	iraq: suicide bombing kills at least 18 in baghdad	iraq: suicide bombing kills at least 22 in baghdad	0.98
5.0	14	albania earthquake kills at least 8	albania earthquake kills at least 13	0.97
100.0	6	iran-iraq earthquake kills more than 300	iran-iraq earthquake kills more than 400	0.98

Table 6: Breakdown of changed nouns. One pair may belong to more than one category

Category	Total	%	Examples (Before, After, #)	Description
Share lemma	132	15.33%	(report, reports, 4); (riots, riot, 3)	The change noun shares the lemma.
Synonym	65	7.54%	(advisor, adviser, 3); (investigate, probe,2)	The change noun is synonym.
Minor	109	12.65%	(protestors, protesters, 5); (reelection, re-election,4)	Only one character is changed (excluding same lemma pairs)
Hyponymn	33	3.83%	(housing, home, 3); (semiconductor, chip, 2), (official, officer, 2); (supporter, believer, 1)	Become more specific
Hypernymn	41	4.76%	(snaps, photos, 5);(budget, plan, 2)	Become more general
Others	481	55.86%	(demo, democrats, 7); (4th, fourth, 4) (live, close, 4); (valuation, value, 3)	All other noun changes
All	861	100%	–	All detected noun substitutions

Table 7: Breakdown of changed verbs. One pair may belong to more than one category

Category	Total	%	Examples (Before, After, #)	Description
Share lemma	135	23.56%	(cancelled, canceled, 7); (shuts, shut, 2)	The change noun shares the lemma.
Synonym	59	10.29%	(rise, climb, 2); (blew, botched, 2)	The change noun is synonym.
Minor	35	6.10%	(eying, eyeing, 2); (targetting, targeting,2)	Only one character is changed (excluding same lemma pairs)
Hyponymn	38	6.63%	(drops, tumbles, 3); (pull, drag, 2) (mulls, considers, 2); (rejects, rebuffs, 2)	Become more specific
Hypernymn	46	8.02%	(passes, advances, 3);(linger, remain, 3) (breaking, damaging, 2); (reveals, shows, 1)	Become more general
Others	260	45.37%	(drop, slip, 3); (reveals, says, 3)	All other noun changes
All	573	100%	–	All detected verb substitutions

Table 8: *Left*: The tracked news categories per agency. *Right*: The headline modification ratio in each news category

Domain	Politics	Business	Entertainment	Technology	Top Stories	Total
BBC	1973	1562	1495	824	0	5854
BuzzFeed	330	92	3928	195	0	4545
CNN	4548	1	1166	629	0	6344
Daily Beast	0	0	0	0	4359	4359
Daily Mail	0	3670	36312	1263	33873	75118
Fox News	6307	108	4273	315	0	11003
Huffington Post	2909	57	1902	7	0	4875
MarketWatch	0	0	0	0	7503	7503
National Review	0	0	0	0	3324	3324
New York Post	0	1458	2018	514	9891	13881
New York Times	2374	2983	3851	538	0	9746
Newsmax	897	25	155	38	0	1115
OAN	2922	3545	528	1328	0	8323
The Blaze	0	0	0	0	4401	4401
The Epoch Times	3179	1077	193	316	0	4765
The Guardian	3456	627	1874	959	0	6916
The Hill	0	0	0	0	5676	5676
Washington Post	2126	2014	1458	478	0	6076
Yahoo News	11074	183653	7	32512	0	227246

Domain	Politics	Business	Entertainment	Technology	Top Stories	Total
BBC	0.3082	0.2855	0.1371	0.0898	0.0000	0.2277
BuzzFeed	0.2576	0.0000	0.4812	0.2154	0.0000	0.4438
CNN	0.2419	0.0000	0.1612	0.0000	0.0000	0.2030
Daily Beast	0.0000	0.0000	0.0000	0.0000	0.1308	0.1308
Daily Mail	0.0000	0.0515	0.0592	0.0451	0.0970	0.0756
Fox News	0.0591	0.0926	0.0314	0.0571	0.0000	0.0486
Huffington Post	0.0399	0.0877	0.0273	0.1429	0.0000	0.0357
MarketWatch	0.0000	0.0000	0.0000	0.0000	0.3453	0.3453
National Review	0.0000	0.0000	0.0000	0.0000	0.0126	0.0126
New York Post	0.0000	0.0638	0.0852	0.0292	0.0650	0.0665
New York Times	0.2784	0.3144	0.0444	0.1375	0.0000	0.1892
Newsmax	0.0268	0.0400	0.0065	0.0000	0.0000	0.0233
OAN	0.0404	0.4886	0.0890	0.2108	0.0000	0.2616
The Blaze	0.0000	0.0000	0.0000	0.0000	0.0400	0.0400
The Epoch Times	0.1787	0.0706	0.0259	0.1203	0.0000	0.1442
The Guardian	0.2922	0.3142	0.0326	0.1022	0.0000	0.1975
The Hill	0.0000	0.0000	0.0000	0.0000	0.0592	0.0592
Washington Post	0.2413	0.1226	0.0501	0.1736	0.0000	0.1508
Yahoo News	0.2390	0.0713	0.0000	0.1073	0.0000	0.0846

Table 9: Examples of altered headlines. The *first* part is the manually selected modifications. The *second* part is sampled from our dataset by change category

Before	After	BERT-F1	Type
One Dose of J.&J. Vaccine Is Ineffective Against Delta, Study Suggests	J.&J. Vaccine May Be Less Effective Against Delta, Study Suggests	0.6830	other
Malawi burns thousands of Covid-19 vaccine doses	Malawi burns thousands of expired AstraZeneca Covid-19 vaccine doses	0.7217	other
Biden to huddle with Senate Democrats on Covid relief ahead of push for passage	Biden urges Senate Democrats to reject poison pills that could sink relief plan ahead of push for passage	0.3777	other
Australia to investigate two deaths for possible links to COVID-19 vaccine	Australia says two deaths not likely to be linked to COVID-19 vaccine	0.7236	citation
Canada adds blood clot warning to AstraZeneca's COVID-19 vaccine	Canada says AstraZeneca COVID-19 vaccine safe, but adds blood clot warning	0.7546	citation
China's Zhurong rover will land on Mars TONIGHT	China's Zhurong rover will land on Mars 'in the next five days	0.7234	other
Nike chief executive says firm is 'of China and for China'	Nike boss defends firm's business in China	0.4559	other
MAGA 2.0	"A new bargain": Biden's 2024 tease bets big on nostalgia	0.0257	emotional
At Least Eight People Died In A School Shooting In Russia	At Least Nine People Died In A School Shooting In Russia	0.9866	paraphrase
India sees cases officially drop below 300,000 a day but now country threatened by killer cyclone	India cases officially drop below 300,000 a day but now country threatened by killer cyclone	0.9034	paraphrase
The Latest: UN: 38,000 Palestinians displaced in Gaza	The Latest: Biden expresses 'support' for Gaza cease-fire	0.4043	dyn update
August 5 at 2PM ET: Join Engine Media CEO and Executive Chairman Fireside Chat	August 5 at 2PM ET: Join Engine Media CEO and Executive Chairman in Fireside Chat	0.9393	dyn update
23 Amazing Jokes From Hot Fuzz	23 Jokes From "Hot Fuzz" That Humans Will Laugh At For The Next 10,000 Years	0.4715	elaboration
Official: Haiti President Jovenel Moïse assassinated at home	Haiti President Jovenel Moïse assassinated at home; Biden calls it 'very worrisome'	0.5179	elaboration
Senator Capito says Republicans plan new U.S. infrastructure offer	U.S. Senate Republicans prepare new infrastructure offer	0.5434	concision
U.S. second-quarter economic growth revised slightly higher; weekly jobless claims rise	U.S. second-quarter growth raised; corporate profits surge	0.5540	concision
Undercover police officers spied on Peter Hain over 25 years	Peter Hain accuses undercover police of lying over reports on apartheid campaign	0.2465	emotional
Suspect arrested following series of Arizona traffic shootings	Arizona gunman goes on traffic shooting rampage, leaving one dead, 12 injured	0.2247	emotional
Raul Castro confirms he's resigning, ending long era in Cuba	Raul Castro resigns as Communist chief, ending era in Cuba	0.7310	neutralize
Nikki Grahame's life will be celebrated in a new Channel 4 documentary four months on from her death	Nikki Grahame to be commemorated in a new Channel 4 documentary	0.6698	neutralize
Fourth stimulus check update: Your next payment could be one of these	Fourth stimulus check? These payments are already in the pipeline	0.4479	foward ref.
12 Movie Opinions You Might Agree With	Here Are 12 More Movie Opinions I Strongly Believe, But Do You Agree With Me?	0.4246	foward ref.
Watch Jeff Bezos' 'Blue Origin' launch into space live today	Everything you need to know as Jeff Bezos' 'Blue Origin' launches into space today	0.6179	personalize
The "Degrassi: TNG" Cast Is Reuniting In Honor Of The Show's 20th Anniversary	The "Degrassi" Cast Is Reuniting And I Am So Excited To Relive My Teen Years	0.4838	personalize
Stimulus Update: States Give Out Thousands of Bonus $1,000 Checks To School Employees	Stimulus Update: States Give Out Thousands of Bonus $1,000 Checks – Will You Get One?	0.7416	personalize
Kendall Jenner Has Some Pretty Strong Thoughts About The Kardashian Curse	Kendall Jenner Talked About The Kardashian Curse And Said "The Men Need To Take Responsibility"	0.5229	citation
Afghan guard killed: Firefight leaves at least one dead and others injured at Kabul airport	'It would be better to die under Taliban rule than face airport crush', say US embassy's 'betrayed' Afghan staff	0.0656	citation
Apple beats sales expectations on iPhone, services, China strength	Apple says chip shortage reaches iPhone, growth forecast slows	0.2611	citation

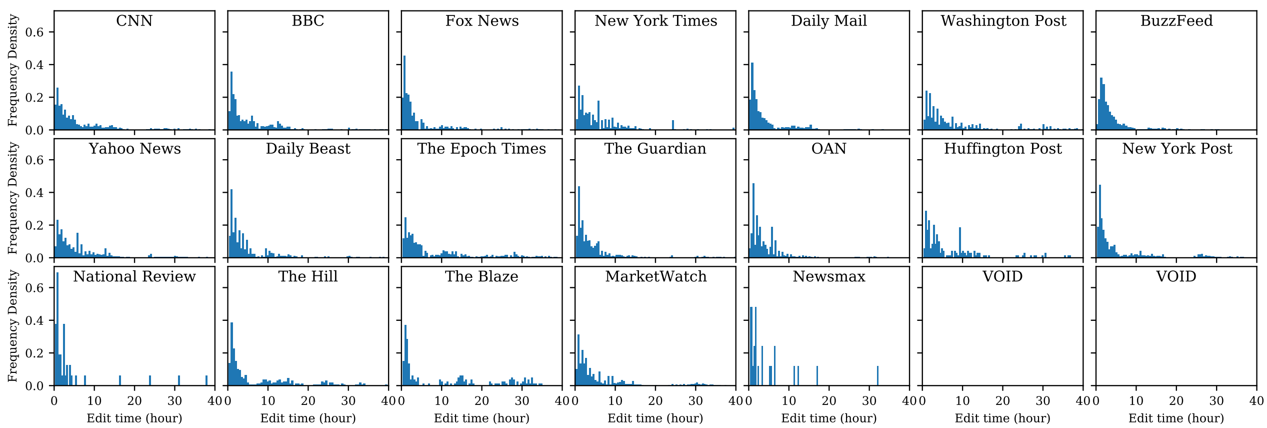

Figure 8: The post-publication edit time per news publisher.