Analyzing Sensationalism in News on Twitter (X): Clickbait Journalism by Legacy vs. Online-Native Outlets and the Consequences for User Engagement

Khawar, Salman; Boukes, M.

DOI
10.1080/21670811.2024.2394764

Publication date
2025

Document Version
Final published version

Published in
Digital Journalism

Volume 13 Number 8 October 2025

Digital Journalism

DIGITAL JOURNALISM

Routledge

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: www.tandfonline.com/journals/rdij20

Analyzing Sensationalism in News on Twitter (*X*): Clickbait Journalism by Legacy vs. Online-Native Outlets and the Consequences for User Engagement

Salman Khawar & Mark Boukes

To cite this article: Salman Khawar & Mark Boukes (2025) Analyzing Sensationalism in News on Twitter (*X*): Clickbait Journalism by Legacy vs. Online-Native Outlets and the Consequences for User Engagement, Digital Journalism, 13:8, 1482-1502, DOI: 10.1080/21670811.2024.2394764

To link to this article: https://doi.org/10.1080/21670811.2024.2394764

DIGITAL JOURNALISM
2025, VOL. 13, NO. 8, 1482–1502
https://doi.org/10.1080/21670811.2024.2394764

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Analyzing Sensationalism in News on Twitter (*X*): Clickbait Journalism by Legacy vs. Online-Native Outlets and the Consequences for User Engagement

Salman Khawar[a] and Mark Boukes[b]

[a]Communication Science, University of Amsterdam, Amsterdam, the Netherlands; [b]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, the Netherlands

ABSTRACT

News consumption patterns have evolved as people increasingly get news through social media. A by-product of this change has been the rise of "clickbaits"—a form of journalism that utilizes sensationalist techniques to lure readers into clicking on links. Although clickbaits have become common, not much is known about the use of this journalistic practice. This study empirically investigates whether online-native news outlets utilize more sensational features than legacy news outlets to promote their news stories on *Twitter* (now *X*). The impact of sensationalism on subsequent user engagement metrics is also tested. First, a manual content analysis of 1,440 tweets by news media organizations was conducted, identifying sensationalism through ten features. Second, an automated content analysis was conducted on a dataset of 25,600 tweets, replicating the manual coding process. The results confirm that online-native outlets (*Huffington Post, BuzzFeed*) utilized more sensationalism than online legacy outlets (*USA Today, L.A. Times*). Furthermore, a positive relationship was found between the number of sensationalist features employed and the favorite count, confirming the effectiveness of clickbaits. The automated content analysis by-and-large corroborated the findings of the manual annotation, and additionally revealed a positive effect of sensationalism on the number of retweets.

KEYWORDS

Sensationalism; journalism; clickbait; Twitter; X; social media news; online news; legacy media; online-native media

Sensationalism in news is a type of journalistic coverage that triggers recipients' reactions, such as attention and emotions, using specific (sensational) production features (Otto, Glogger, and Boukes 2016). Initially, sensationalism was often conflated with certain topics or news categories, such as crime, violence, natural disasters, accidents or human-interest (Adams 1978). Later, a clearer distinction started being drawn between the sensational *content* and the sensational *form* of news stories; content referring to the choice of topic and form referring to the stylistic or production features being used (Molek-Kozakowska 2013).

CONTACT Mark Boukes markboukes@gmail.com

Supplemental data for this article can be accessed online at https://doi.org/10.1080/21670811.2024.2394764.

As the younger generations embrace the more participatory and personalized news offered by social media platforms, the news industry consequently has gone through a "digital transformation", moving from broadcast and print to digital (Newman et al. 2023, 5). *Twitter* (nowadays: *X*), in particular being a popular news-oriented platform, is the focus of this research that investigates how journalistic organizations (i.e., legacy vs. online-native) differentially use sensationalism to promote their news stories on this platform and how this, subsequently, influences user engagement.

Twitter has positioned itself as a social media platform used for event-tracking—rather than for sharing personal content or ambient chatting, such as *Facebook* or *Instagram* (Safori 2019). The appeal and success of *Twitter* lies in its simplicity, primarily designed for mobile phone users who can tweet anytime anywhere, making it a powerful distribution channel for breaking news (Ju, Jeong, and Chyi 2013). It is primarily a textual platform, rooted in SMS practices, that privileges short, concise, and to-the-point messages (Hermida and Mellado 2020). *Twitter* has a high percentage of users who use the platform as a regular source of news (55%) compared to other social media, such as *Facebook* (31%; Walker and Matsa 2021). In fact, journalistic outlets have had four times more followers on *Twitter* than on *Facebook*; thereby driving more traffic to news websites relative to other platforms with greater reach (Ju, Jeong, and Chyi 2013).

The affordances of *Twitter* that make it an often-used source for news, also make it conducive to a sensationalist form of news (Pelau et al. 2023). Research suggests that information diffuses faster, farther, and deeper on *Twitter* (Vosoughi, Roy, and Aral 2018), which also includes false claims. Sensationalist news can, thereby, lead to adverse consequences for users such as increasing skepticism (Staender et al. 2021), social media fatigue, and discontinuous behavior (Pelau et al. 2023). Such negative perceptions surrounding news can eventually lower individuals' trust in news media and weaken their ability to distinguish between real and fake news (Van Duyn and Collier 2018). Considering its importance for the industry and audiences, the journalistic use of sensationalism on *Twitter* is investigated in this study.

In an online news environment where media outlets are vying for users' attentions, who are spoilt for choice and have limited attention spans (Subramanian 2017), the attractiveness of news has arguably become more important than the source's credibility (Pelau et al. 2023; Chakraborty et al. 2017). In this context, sensationalism has become increasingly relevant and is widely employed in journalistic practice, in the form of clickbaits. Clickbait journalism is often held responsible for misleading readers, propagating fake news, and violating journalistic codes of ethics (Scott 2021; Potthast et al. 2018). Notwithstanding this negative connotation, a successful social media clickbait headline manages to: (i) attract users' attention while browsing and (ii) arouse curiosity inducing them to click on a link (Scott 2021).

News outlets employ various sensationalist features on social media to lure readers into clicking on links to news items on their websites: a process called "clickbaiting" (Chakraborty et al. 2017). Due to a lack of consensus on what constitutes sensationalism on social media—most existing research still pertains to traditional audio-visual media—the current study operationalizes the sensationalism construct for the digital context through ten commonly used sensationalist features based on linguistic properties of social media posts.

A total of 1,440 *tweets* were analyzed using manual content analysis. Tweets were extracted from four news outlets' accounts (two online-native: *BuzzFeed*, *Huffington Post*; and two legacy: *L.A. Times*, *USA Today*) posted between September 9, 2015 and December 2, 2021, to test how sensationalism is utilized differently by legacy versus online-native outlets. Moreover, the relationship between the use of sensationalist features in tweets and subsequent user engagement (measured through the metrics of *favorite* count and *retweets*) was investigated to explore clickbait's impact. On a larger sample of 25,600 tweets, an automated content analysis was conducted. The same ten sensationalist features were coded and the results were validated against the manually annotated tweets. Thereby, we provide a conceptual replication of the manual analysis and provide a method to scale up this study's design for larger samples in future research.

Theoretical Background

Sensationalism in the Digital Context

Sensationalism is not something new but can be traced back to the nineteenth century with the emergence of the "penny press" (Bird 2009). These newspapers replaced substance with sensation and mostly covered stories involving sports, crime, tragedy, adventure and gossip (Friedman 2005). Publications such as William Randolph Hearst's *New York Journal* used "bold and brash headlines, hyperbolic rhetoric, enticing imagery and other characteristics associated with sensationalism", almost similar to what *BuzzFeed* does today (Berman 2021, 4110). A comparative analysis between these early tabloid newspapers and online news outlets suggests that both used sensationalism to adapt to the highly concentrated "marketplaces of attention" in which they operated—the bustling New York streets vis-à-vis the modern social media news feeds (Berman 2021).

Sensationalist journalism in the digital realm *also* seeks to deliberately engage audiences through simplistic, attractive content that appeals to the audience's emotions or piques their interest (Brown and Sinta 2016). Nowadays, almost all news media organizations promote their news stories on *Twitter/X* through catchy sensational posts with the goal to attract more users through clickbaiting (i.e., a sensationalized headline that tempts the reader to click on the link to a news story (Brown et al. 2018). To achieve higher visibility on such social media platforms, news outlets need to abide by the *social media logic*—the norms, strategies, mechanisms and economies underpinning these spaces (Van Dijck and Poell 2013). If this is done in a way that is compatible with the technological affordances and communicative rules of the respective platform, news outlets are rewarded by algorithms (Klein et al. 2023), which means greater reach for their posts. Hence, sensationalism on social media, arguably, differs from sensationalism in the traditional news media.

Whereas most legacy news media's revenue still comes from subscriptions, the online news media operate differently as they are mostly competing for users' attention, who have a plethora of options and very limited attention spans (Subramanian 2017). The revenue model for most online media outlets, which provide free news, is reliant on advertisements on their webpages and the most important metric for

that is the number of "clicks." Moreover, some media outlets even pay their journalists bonuses for the clicks their stories generate (Murtha 2015), which incentivizes the production of sensational news headlines. It is argued that the attractiveness of news has consequently become more important than the credibility of the source (Pelau et al. 2023; Chakraborty et al. 2017).

Earlier studies on sensationalism, focusing on traditional media, had a tendency to operationalize sensationalism dichotomously on *content* as either being entertaining and attention-grabbing (crime, violence, natural disasters) or "proper" content that would enhance political and social knowledge (Arbaoui, De Swert, and van der Brug 2020). Later studies started focusing more on the way that news stories are packaged or presented, also known as journalistic *form,* involving narrative style and production elements within news stories (Uribe and Gunter 2007). There is, however, still a need to define what sensationalism encompasses on social media. Well-defined stylistic production features for televised news, such as specific camera angles/maneuvers and decorative effects, are not suitable indicators for sensationalism on social media, which means that alternative indicators of sensationalism need to be developed for news online.

Sensationalist Features in Social Media News

This study analyzes news outlets' social media posts—specifically: tweets—functioning as online "teasers" (Potthast et al. 2018). These posts promote news items published by news outlets to draw users to their websites where eventually advertising revenue can be generated. Thereby, these tweets are somewhat similar to headlines in print media, which are considered the main "hook" for readers (Molek-Kozakowska 2013).

Previous work that analyzed sensationalism in online news has been guided by their methodological approaches. Studies relying on manual coding have evaluated the sensationalist treatment of news more in-depth, for example, in terms of news values, topics and potential to evoke emotions and capture attention (Brown et al. 2018; García-Perdomo et al. 2017). Research using computational methods has analyzed larger amounts of data, inductively extracting sensationalist features by comparing linguistic differences between clickbait and non-clickbait datasets (Chakraborty et al. 2017; Biyani, Tsioutsiouliklis, and Blackmer 2016) or using lexica- and rule-based approaches (Jung et al. 2022). Validated methods for identifying sensationalism have looked at the phenomenon at various levels ranging from individual word (lexical) choices, language and grammatical (syntactic) structures to the genre/subject matter (Chen, Conroy, and Rubin 2015). The current study adopts a hybrid approach using content-based sensational features appearing in tweets that are directly measurable (Firdaus, Ding, and Sadeghian 2021). Moreover, the current research uses both manual and automated content analysis, as to leverage the qualities of both approaches; and also replicating the first findings with a larger sample through computational methods.

There is a lack of agreement, in academia and industry, on what constitutes sensationalism in online news, though. Molina et al. (2021) identified seven recurrent characteristics of clickbait headlines including questions, lists, demonstrative adjectives and superlatives. Kemm (2022) identified twelve linguistic and typological

features found in *YouTube* video titles. We expand upon such previous efforts by operationalizing sensationalism more holistically, through ten theoretically informed features based on the syntactic structure, lexical choices, informality, and media used in tweets. The ten features used to identify sensationalism in this research are explained below.

1. *Hyperbole,* or extravagant language, is often used to drive referral traffic on social media. The use of superlatives can be seen as a structural response to the competitive environment of the news feed, as doing so raises expectations about the contents of the article being promoted (Berman 2021). Using these sensationalist features, such as intensifiers, can impact the audience's evaluation of the news article as well as its perceived newsworthiness (Burgers and de Graaf 2013).
2. *Forward referencing* is a feature used in online news that aims to create curiosity or an information gap among readers as it refers to a discourse segment. This information can only be viewed by accessing the full-text by clicking on the accompanying link (Blom and Hansen 2015). A clear example is news posts starting with "This is why ….".
3. *Listicles* are a new media phenomenon used to structure headlines comprising of lists or rankings with concise and easily understood details about a topic for readers (Sadri 2019); e.g., "17 Real-Life Secrets About ….". Listicles have a homogeneous structure with 85% of the headlines starting with a cardinal number, representing the number of items listed (Potthast et al. 2016). Research suggests that 13% of all internet users consume news in the form of lists (McCartney 2016).
4. Another way news organizations attract attention is by utilizing the interrogative structure; presenting information in the form of a *question*, giving the impression that it was previously hidden (Molek-Kozakowska 2013). This also creates a curiosity gap and forces the reader to click on the link to find out the answer to the question. Example: "GOP operatives want Kanye West on the Presidential ballot. Will it really matter?".
5. The overuse of capitalization or *all-caps words* adds emphasis and attracts users' attention. It is often used to distinguish between clickbait and non-clickbait headlines, for example by Biyani, Tsioutsiouliklis, and Blackmer (2016).
6. *Emojis* are Unicode graphic symbols used to express emotions, ideas and concepts (Novak et al. 2015). The use of *emojis* has been studied in the context of news dissemination on social networking sites and has a clear potential to draw audience attention (Beyersmann, Wegener, and Kemp 2023).
7. *Hashtags* are not only a commonly used informal feature in posts, but also help users to track and follow information on specific topics on *Twitter* (Firdaus, Ding, and Sadeghian 2021). They are proven to have an impact on user engagement (Orellana-Rodriguez, Greene, and Keane 2017).
8. Similarly, *informal punctuation* is used on social media to highlight and emphasize something that is considered important by the author; for instance, "!!!", "?!?". It adds subjectivity and allows to exclaim or question something more vividly (Welbers and Opgenhaffen 2018).

9. *Internet slang* is another feature commonly used in clickbait headlines to grab readers' attention (Chakraborty et al. 2016). Slang is informal usage of vocabulary and idiom that is characteristically more metaphorical, playful, elliptical, vivid and ephemeral than ordinary language (Puspitorini 2020).
10. *Twitter* enables its users to include *media* in the form of photos, videos or animated GIFS in their posts. Such imagery can be used as potent weapon to differentiate content, by choosing images that elicit an immediate reaction over ones that provide deeper context to the news article (Berman 2021). Literature suggests that most (legacy) news organizations still lag behind when it comes to using such media features (Armstrong and Gao 2010).

Legacy versus Online-Native News Outlets

To understand the latest practices of journalists and evolution of journalism, it is imperative that the new players in the media landscape are also investigated. Especially, because some of these have become leaders of web traffic and social media sharing in recent years (Brown and Sinta 2016). Yet, online news outlets are often still overlooked in research due to the misconception that their websites do not have real news or are just proliferated with "funny cat videos" (Canter 2018).

Various terms and labels have been used for news organizations that emerged on the internet, such as "new media," "native online media," "pure online media" and "digitally-born" (Brown et al. 2018). This research uses the term "online-native" for those news outlets that were created for online audiences without an existing prior print or TV counterpart. On the other hand, the label "legacy" is used for traditional news outlets that originally operated as physical newspapers, but now have also branched out to online and social media. The hypothesis that we test is whether the legacy media have adopted the sensationalist features to a different degree than the online-native outlets.

The distinction between online-native and legacy news outlets is particularly interesting considering the differences in their journalistic offerings, the audiences they cater to, and their business models (Arrese and Kaufmann 2016). Legacy media tends to abide by the norms and practices of traditional journalism, whereas journalists working in online-native news have more "flexible boundaries" for journalistic norms (Brown et al. 2018). Legacy brands seek to integrate their traditional readers with online readers; trying to find a balance between quantity and quality in such a way that charging for content is still feasible. Online-native brands on the other hand create their own audiences, with a preference of quantity over quality (i.e., potentially more clicks), relying on a business model based on free news (Arrese and Kaufmann 2016). For publishers of online news to be financially viable, their content is expected to "pay for itself" by attracting enough visitors on their pages, through which they earn advertisement revenue (Potthast et al. 2018).

Online-native sites, such as *BuzzFeed* or *Huffington Post*, emphasize shareability as the foremost virtue, often ignoring the veracity of the story; in other words "shareability without responsibility" (Riordon 2014, 50). Legacy outlets, on the contrary, have robust systems of verifying information before being published in digital forms (Riordon 2014). Legacy media outlets face additional pressures of delivering quality

content and protecting their "brand" because audiences have higher expectations from them, making it difficult to compete with new players in online news (Siegert et al. 2016, 399).

Rony, Hassan, and Yousuf (2017) discovered that online-native outlets, or "unreliable media" (235) as they term it, use more clickbait headlines than mainstream media. Skovsgaard (2014) discovered that journalists who use a popularized, personalized and sensationalist journalistic style focus more on profits—weakly adhering to objectivity and public service obligation—compared to traditional journalists. Hence, we expect that online-native news outlets are more likely to use sensational features than the legacy outlets:

H_1: *Online-native news organizations include more sensationalist features in their tweets than legacy news organizations.*

Sensationalism and User Engagement

Researchers and practitioners have been trying to examine reasons behind higher diffusion and popularity rates of certain social media news posts, measured through engagement metrics (Jung et al. 2022; Brown et al. 2018; García-Perdomo et al. 2017). The Elaboration Likelihood Model (ELM; Petty and Cacioppo 1986) can be used to understand how tweets containing sensationalist features might lead to more engagement. According to the ELM, users evaluate information through two paths, namely the central and peripheral routes. Generally, information evaluated through the central route requires more cognitive effort, whereas information processed through the peripheral route is based on heuristic cues (Pelau et al. 2023). This is grounded in the idea that social media users are cognitive misers as they operate in already overloaded online information environments, and relying on heuristics allows them to minimize their analytical effort (Sui et al. 2023; Metzger, Flanagin, and Medders 2010). One could argue that sensationalist features play the crucial role of such heuristic cues; thus, determining whether one wants to engage with a tweet or not.

Relatedly, prior research has tried to explain the psychology behind the effectiveness of sensationalist linguistic strategies employed by clickbaits through Loewenstein's (2008) "information gap theory of curiosity". Sensationalist posts aim to trigger curiosity by providing enough information, beyond the users' current knowledge of a topic, to pique curiosity but not sufficient to fulfill it; hence, creating a knowledge gap (Molina et al. 2021; Scott 2021). The user—in order to fill this gap—feels the need to click on the accompanying news item.

A review of prior literature on the impact of sensationalism on user engagement reveals contradictory findings. Most research suggests that clickbait content receives more engagement, in the form of reactions/shares/comments (Lamot, Cools, and Gevers 2024; Rony, Hassan, and Yousuf 2017; Chakraborty et al. 2017). However, Scacco and Muddiman (2019) in their experimental study found a "backfire effect" by which lower perceptions of information adequacy for such headlines resulted in less engagement. The main reasons behind this contradiction are likely differences in the operationalization of sensationalism (selected features) and the methodological

approaches adopted (Molina et al. 2021). Since the current research uses a combination of linguistic features and a methodological approach more in line with the first strain of research, we expect that sensationalist features will positively influence engagement.

Although social media platforms are constantly updating their features, we focus on two basic engagement metrics common to most platforms: favorite count and retweets. These metrics can be seen as "affordances" provided to users that determine how they interact with each other and with the technology (Brown et al. 2022). Liking a post indicates that users agree with or appreciate its contents. Hence, through the "like" button social networking sites provide the affordance of social validation (Brown et al. 2022), which may encourage other users to also read their news. The "retweet" button enables users to actively (re)share other-sourced tweets, which can be seen by their followers immediately. Thus, retweeting provides a powerful way for users to construct online identities, attract more followers, and establish relationships (Shi, Lai, and Chen 2023); simultaneously increasing the reach and potential audience of online news outlets.

Considering the alleged popularity of sensational elements in social media posts, the congruence with heuristic information processing in the online domain, and the potential to create a curiosity gap, we predict the following:

> H_2: *Tweets of news organizations with more sensationalist features generate more engagement—in the form of (a) more likes and (b) more retweets—than tweets of news organizations with fewer sensationalist features.*

Methodology

Choice of News Outlets

BuzzFeed and *Huffington Post* were chosen as the two online-native outlets in this research. These websites have been used extensively in the study of sensationalism and clickbaits (Berman 2021; Scott 2021; Chakraborty et al. 2017; Potthast et al. 2018; Brown et al. 2018; Brown and Sinta 2016). *BuzzFeed* is a popular online publisher particularly prominent amongst millennials (Cowley 2017). It has one of the largest digital audiences in the U.S and is a pioneer of sensationalist strategies associated with viral news (Berman 2021). The *Huffington Post* is another prominent internet news website, blog and news aggregator; it became the flag-bearer of online journalism, when this new field was undergoing the process of institutionalization (Neilson 2012).

Two main factors were considered in the selection of legacy news outlets (Watson 2023). The first was a relatively high *content comparability* with the online-native outlets in terms of news topics covered. Previous research indicates that entertainment-oriented news topics are generally covered in a more sensationalist manner compared to politics-oriented news topics (Scott 2021; Rony, Hassan, and Yousuf 2017; Curran et al. 2010). To avoid such bias towards any one topic or style of reporting in the sample the criterion was for news outlets to cover a range of topics and have separate *Twitter* accounts/sub-channels with substantial followers

for politics and entertainment news topics; so, we could control for this in the analysis.

The second factor was *comparability in audience composition* of the legacy and online-native news outlets, because popular news media can widely differ from quality news media in their coverage of the same issues (Boukes and Vliegenthart 2017). For a fair comparison, we considered middle-market newspapers that are neither upmarket (primarily hard news) nor completely downmarket (primarily sensationalist). *USA Today* and *L.A. Times* were natural choices as legacy news outlets because they cater to a similar audience with a similar type of journalism as the selected online-native outlets. *USA Today* has been an early adopter of technology and even received opposition from critics and journalists for its stylistic innovations (Bird 2009). Similarly, *L.A. Times* has over the years made attempts to engage younger readers who use social networking sites, such as introducing a multimedia editorial product known as "Brand X" in 2009 (Stanton 2009). In terms of brand trust scores, the selected news outlets share similar scores, which are considerably lower than the quality news media (Newman et al. 2023).

Data Processing and Sampling

Data were collected through *Twitter*'s API using the Python library *Tweepy* in the first week of December, 2021.[1] Due to API restrictions, a maximum number of 3,250 most recent tweets per account were extracted from the following accounts: @BuzzFeedPol, @BuzzFeedSports, @BuzzFeedEnt, HuffPostPolitics, @HuffPostSports, @HuffPostLife, @usatodayDC, @usatodaysports, @usatodaylife, @latimespolitics, @latimessports, @latimesent. To ensure that these handles were owned and operated by the selected news outlets, the presence of the blue checkmark (tick) was a prerequisite, which was *Twitter*'s verification mechanism at the moment of data collection. The timeframe for the extracted tweets was from September 9, 2015, to December 2, 2021. Tweets were predominantly in English except for *L.A. Times Entertainment*, which also contained a significant number of posts in Spanish. The Python library *DeepL* was used to translate these tweets into English.

As two of the three handles (entertainment/lifestyle and sports) selected for each news outlet were entertainment-oriented, stratified samples were drawn to ensure parity with the politics-oriented posts and to prevent "topic bias" (Potthast et al. 2018, 1503). This approach to sampling was deemed necessary to ensure that the samples drawn were more inclusive and representative—containing a range of topics and not just biased towards any particular topic—consequently leading to higher validity of results. For greater granularity, the topics of individual tweets were coded as well (into ten topic categories; Krippendorff's $\alpha = 0.71$) and included in the statistical analyses (for manual content analysis) as a control variable.

From the raw data of 38,838 posts, random samples for automated content analysis ($n = 25{,}600$), manual content analysis ($n = 1{,}440$), and intercoder reliability testing ($n = 144$) were drawn.[2] The samples constituted of 50% posts taken from the politics sub-channels and the other 50% from the entertainment-oriented sub-channels, split evenly by sports and entertainment/lifestyle, to have a fair balance (see Table A1 in Online Appendix).

Measurements: Manual Content Analysis

The codebook was prepared with instructions and examples illustrating how sensationalist features were to be identified.[3] Binary variables were created for the ten sensationalist features (0: not present *vs*. 1: present); Table 1 displays each coding instruction. Binary variables were also created for individual topic categories, which were then recoded into politics-oriented topics (politics/government/military, international affairs, economy/business, health/education and science/technology) and entertainment-oriented topics (entertainment/celebrity news, lifestyle and society, sports, human-interest story, crime/law and order).

Tweets were manually coded in *Qualtrics* by the manuscript's first author (coder 1). Prior to coding the main sample, a second coder (coder 2) independently coded the ICR sample.[4] The intercoder agreement for all variables exceeded the minimum acceptable score (Krippendorff's $\alpha > .67$). The use of media was the most popular feature; 20.21% of tweets were accompanied by either a photo, video or GIF. The second most prevalent feature was the presence of hyperbolic words (10.01%), followed by hashtag/s (8.47%), and internet slang (5.83%).

The maximum number of sensationalist features found in a single tweet was five (only in 0.21% of cases). A majority of sensationalist tweets, however, contained no (55.97%) or one sensationalist feature (28.61%); 10.90% of tweets had two sensationalist features, 3.61% had three, and 0.69% of tweets contained four sensationalist features (Figure 1).

Table 1. Manual content analysis: Features, coding instructions, and inter-coder reliability.[a]

Sensationalist feature	Coding instructions summary	Presence (% of tweets)	Krippendorff's α
Hyperbolic words/phrases	Hyperbole can be identified through words/phrases such as "amazing", "greatest", "best ever", "mind blowing", "epic".	10.01%	0.69
Forward referencing	Can usually be identified if sentences in tweets start with demonstrative pronouns/adjectives (this, that, these, those) or adverbs (here, there).	4.17%	0.76
Listicle	Can usually be identified when tweet text contains a number referencing something in the accompanying news item.	0.90%	1.00
Question/s	Tweet has an interrogative structure i.e., contains question mark ("?").	5.56%	0.85
All Caps word/s	Tweet text contains one or more words in "all caps," except for names/acronyms such as "COVID-19", "GOP", "DOJ" etc.	3.06%	1.00
Emoji/s	Tweet contains one or more emoji/s.	4.24%	1.00
Hashtag/s	Tweet text contains one or more hashtags ("#").	8.47%	0.95
Informal punctuation	Tweet text contains informal punctuation patterns such as "!?," "…," "***," "!!!."	2.57%	0.91
Slang	Tweet text contains internet slang words such as "OMG", "AF", "BFF" in either lower or upper case.	5.83%	0.74
Media (photo/video/GIF)	Tweet is accompanied by a photo (code 1), video (code 2), or animated GIF (code 3). if no media then code 0.	20.21%	0.91

[a]Jupyter notebook with Python code for manual content analysis can be accessed at https://bit.ly/Content-Analysis-Results

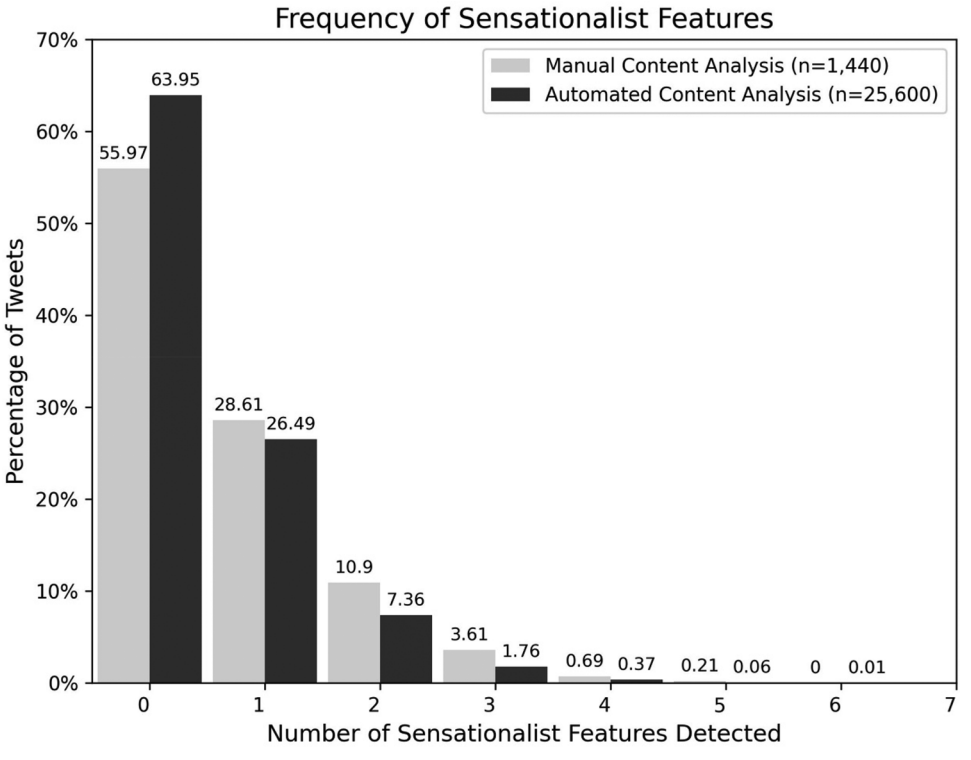

Figure 1. Percentage distribution of tweets by number of sensationalist features.

Measurements: Automated Content Analysis

Computational text analysis methods involve a range of tools from extracting specific contents using simple keywords or formatting rules to statistically complex software solutions (Baden et al. 2021). Natural language processing (part-of-speech tagging, regular expressions) and lexicon-based approaches were employed in the current research. Such techniques can be classified as deductive or top-down approaches (Boumans and Trilling 2016).

For the automated content analysis of the larger sample of 25,600 tweets, *Python* was used to detect the sensationalist features. At least one sensationalist feature was present in 36.05% of all tweets according to this analysis. The approaches used and detection rates for all features are delineated in Table 2.

Results

Findings of Manual Content Analysis: Hypothesis Testing

News Outlet Type and the Use of Sensationalist Features (H_1)

Two regression tests were run to test H_1. First, a logistic regression analysis to determine the likelihood of *presence* of sensationalist feature/s (dichotomous dependent variable: one or more sensationalist feature/s present = 1, no sensationalist features = 0); and, second, a Poisson regression to predict the *number* of sensationalist features

in tweets. The main independent variable was outlet type, with channel type, topic and tweet length as covariates.

The logistic regression results suggest that online-native news outlets are more likely to use sensationalist features in tweets compared to legacy news outlets ($b = 0.79$, $OR = 2.21$, $p < .001$) (see Table 3). Concretely, this means that at least one sensationalist feature was found in 51.81% of the tweets by online-native media versus 36.29% of tweets by the legacy media (see Figure 2).

The Poisson regression results additionally found that tweets published by online-native news outlets are likely to have more sensationalist features compared to tweets published by legacy news outlets ($b = 0.46$, $SE = 0.07$, $p < .001$). Thus, the results provide robust evidence supporting H_1; online-native outlets do utilize *more* sensationalism in their tweets compared to legacy outlets.

Sensationalist Features and Engagement Metrics (H_2)

Poisson regression models were run with engagement metrics (likes and retweets) as dependent variables to test the second hypothesis; the number of sensationalist features present in a tweet functioned as the main independent variable and the same control variables were used as for H_1. The results suggest that a greater number of sensationalist features in tweets leads to more likes ($b = 0.31$, $SE = 0.01$, $p < .001$) (see Table 3). A significant relationship was not found for the number of retweets ($b = 0.00$, $SE = 0.01$, $p = 0.773$).

To verify the robustness of this finding, another Poisson regression analysis was run with the (dichotomous) variable for the presence of sensationalism as the

Table 2. Automated content analysis: approaches and feature detection.[b]

Sensationalist Feature	Approach used in Python	Detection Rate (%)
Hyperbolic words/phrases	Part-of-speech tagging (PoS; see Biyani, Tsioutsiouliklis, and Blackmer 2016) and the *Downworthy* browser plugin corpus (Jung et al. 2022; Gianotto 2016) – were used to detect hyperbole.	2.66%
Forward referencing	A list of words (demonstrative pronouns/adjectives), associated with this feature, was compiled and used to check for relevant usage of these words.	4.43%
Listicles	Regular expressions were used to detect tweets starting with relevant digits.	0.77%
Questions	Regular expressions were used to identify the presence of "?" in text.	5.34%
All-caps words	Regular expressions were used. Proper nouns/ acronyms were excluded through PoS to avoid wrongly coding posts.	0.88%
Emojis	The *emoji* Python library was used to detect this feature.	2.89%
Hashtags	A function was used to detect the presence of "#" in posts.	8.90%
Informal punctuation	Regular expressions were used to detect irregular or informal punctuation patterns.	2.55%
Slang	Two dictionaries (Mbaye 2019; Walter 2019), containing commonly used slang and abbreviations in tweets, were used.	1.79%
Media (photo/video/GIF)	Data extracted from the metadata (*extended_entities* object) provided by the *Twitter* API.	18.10%

[b]Jupyter notebook with Python code for automated content analysis can be accessed at https://bit.ly/ACA-Results

Table 3. Regression results for manual content analysis ($n = 1{,}440$).

	Dependent variables			
	H_1		H_2	
Independent variables	Presence of sensationalism (binary)	Number of sensationalist features (count)	Favorite count	Retweet count
Intercept	−1.92** (0.02)	−1.98** (0.12)	1.24** (0.03)	1.39** (0.03)
Number of sensationalist features (count)			0.31** (0.01)	0.00 (0.01)
Outlet Type: online- native (1) vs. legacy (0)	0.79** (0.12)	0.46** (0.07)	0.55** (0.02)	1.16** (0.02)
Topic: Entertainment-oriented (1) vs. Politics-oriented (0)	0.59* (0.20)	0.54** (0.14)	0.94** (0.02)	0.32** (0.03)
Channel Type: entertainment/ lifestyle (1) vs. politics (0)	1.29** (0.20)	0.97** (0.13)	−1.01** (0.02)	1.27** (0.03)
Tweet length	0.00 (0.00)	0.00** (0.00)	0.01** (0.00)	0.00** (0.00)
Regression model:	Logistic	Poisson	Poisson	Poisson

Standard errors (*SE*) reported in parentheses. *indicates *p*-values < 0.05, **indicates *p*-values <0.001.

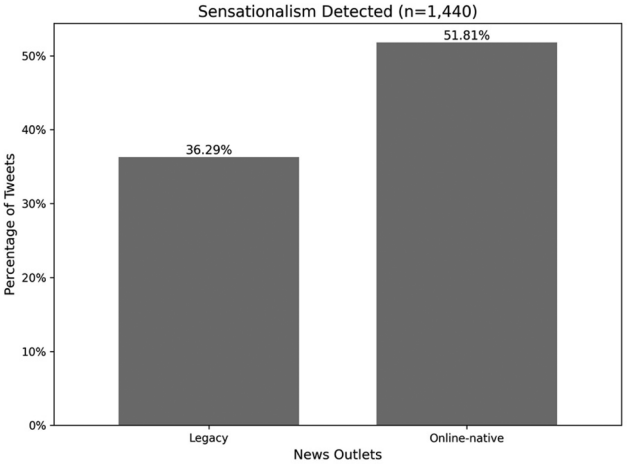

Figure 2. Presence of sensationalism by type of news outlet (manual content analysis).

independent variable. The results confirmed that tweets containing one or more sensationalist features received significantly more likes ($b = 0.37$, $SE = 0.02$, $p < .001$), but lesser retweets ($b = -0.32$, $SE = 0.02$, $p < .001$), than tweets without sensationalist features (see Table C2 in Online Appendix). Considering the aforementioned H_2 can only be partly confirmed.

Next, we explored *which* of the sensational features specifically impacted the engagement metrics (see Table C3 in Online Appendix). For the favorite count, positive statistically significant relationships were found with the following sensational features: slang ($b = 0.57$, $SE = 0.03$, $p < .001$), emojis ($b = 1.17$, $SE = 0.03$, $p < .001$), questions ($b = 1.12$, $SE = 0.02$, $p < .001$), hashtags ($b = 0.30$, $SE = 0.03$, $p < .001$), and media features ($b = 0.11$, $SE = 0.02$, $p < .001$).

For the retweet count, positive statistically significant relationships with the following features were established: slang ($b = 0.33$, $SE = 0.02$, $p < .001$), hashtags ($b = 0.32$, $SE = 0.02$, $p < .001$), informal punctuation ($b = 0.21$, $SE = 0.03$, $p < .001$), and the all-caps

words ($b = 0.83$, $SE = 0.02$, $p < .001$) features. Slang language and hashtags, thus, positively influenced both types of engagement in a positive direction. Interestingly, hyperbolic words/phrases, listicles, and forward referencing features had a *negative* (and statistically significant) impact on both the likes and retweets.

Findings of the Automated Content Analysis: Validation and Hypothesis Testing

When dictionaries or classifiers are used in a context other than what they were developed for, their results should be validated against manual coding (Boukes et al., 2020; Song et al. 2020). In this research, the manually annotated sample of 1,440 tweets was used to validate the automated techniques used. The validation/evaluation metrics for each feature are outlined in Table 4, which found positive correlations for all features and mixed results for their reliability.

Listicles, questions, emojis, informal punctuation, and hashtags were well-retrieved by the automated measurements. Features that were relatively difficult to detect were hyperbolic words/phrases and internet slang. The context-specific usage of these two features in our sample reduced the effectiveness of computational methods employed. Hence, the recall for these features was relatively lower; potential consequences are elaborated in the Discussion section.

As shown in Table 5, the results for the hypothesis testing by-and-large confirmed the results of the manual content analysis. For H_1, we again find that the online-native

Table 4. Validation of automated content analysis.[a]

Feature	Precision	Recall	F1 score	Correlation (rs)	Krippendorff's α
Hyperbolic	.81	.23	.36	.32	.34
Forward referencing	.61	.73	.67	.65	.65
Listicle	.90	.69	.78	.79	.78
Question	.96	.97	.97	.97	.97
All-caps	.71	.45	.56	.56	.54
Emoji	.98	.74	.84	.84	.84
Hashtag	.98	.99	.99	.99	.99
Informal Punctuation	.86	.68	.76	.76	.75
Slang	.55	.20	.30	.31	.27

[a]Validation scores for the media feature were excluded as its detection involved extracting metadata of tweets.

Table 5. Regression results for automated content analysis ($n = 25{,}600$).

	Dependent variables			
	H_1		H_2	
Independent variables	Presence of sensationalism (binary)	Number of sensationalist features (count)	Favorite count	Retweet count
Intercept	−2.13** (0.05)	−2.01** (0.03)	2.10** (0.01)	1.92** (0.01)
Number of sensationalist features (count)			0.16** (0.00)	0.32** (0.00)
Outlet Type: online- native (1) vs. legacy (0)	1.02** (0.03)	0.59** (0.02)	0.96** (0.00)	1.32** (0.00)
Channel Type: entertainment/lifestyle (1) vs. politics (0)	1.63** (0.03)	1.28** (0.02)	−0.12** (0.00)	1.81** (0.00)
Tweet length	0.00** (0.00)	0.00** (0.00)	0.00** (0.00)	−0.01** (0.00)
Regression model:	Logistic	Poisson	Poisson	Poisson

Standard errors (*SE*) reported in parentheses. * indicates *p*-values < 0.05, ** indicates *p*-values <0.001.

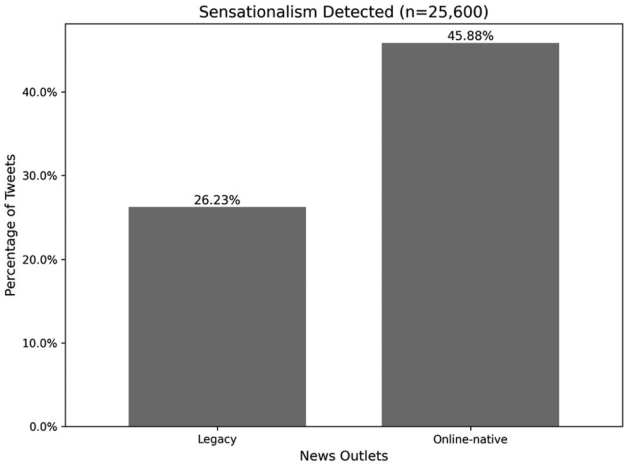

Figure 3. Presence of sensationalism by type of news outlet (automated content analysis).

outlets used more sensationalism in their tweets than the legacy outlets (see Figure 3). For H_2, the results slightly differed as not only a positive statistically significant relationship was found between sensationalism in tweets and favorite count, but in this case the pattern was also confirmed for retweets; thus, actually finding stronger support for the second hypothesis than the manual content analysis. (For detailed results see Tables B1, C1, C2, C3 in the online Appendix).

Discussion

Contributions to the Literature

Sensationalism in social media news is an under-researched topic, especially considering its relevance and impact on people's daily lives as well as the journalism industry. Previous research on online sensationalism has mostly been motivated by the aim of developing automatic clickbait detection mechanisms (Potthast et al. 2018; Chakraborty et al. 2016; Chen, Conroy, and Rubin 2015), or in studying misinformation (Staender et al. 2021). Little work has considered clickbait from a journalistic, linguistic or pragmatic perspective (Scott 2021). The current study contributes to the literature by operationalizing sensationalism through nine linguistic features (and one media feature). Furthermore, it specifically looks at the differential usage of sensationalism, between online-native and legacy news outlets on *Twitter,* and its impact on user engagement.

Our findings demonstrate that online-native news outlets use sensationalism more frequently and to a higher degree (number of sensationalist features) than legacy news outlets. A case study conducted by Ryfe (2009) already illustrated that (legacy) journalists are resistant to change. Journalists' routines are deeply rooted in an embedded culture of professionalism associated with traditional journalistic norms and practices; this culture might also hamper the uptake of sensationalism in the already established newsrooms of online legacy media. Indeed, our findings suggest that in new environments, such as newsrooms of online-native outlets, existing journalistic

cultures may be less dominant; thereby, giving more space to experiment with journalistic formats.

Previous research has been divided about what elements constitute sensationalism. For example, Blom and Hansen (2015) focused on forward referencing, Biyani, Tsioutsiouliklis, and Blackmer (2016) investigated informality, and Sadri (2019) studied listicles. The current research attempts to synthesize the various perspectives to comprehensively operationalize sensationalism. For this purpose, a codebook has been created, which will also guide future researchers interested in studying sensationalism or clickbait journalism.

Methodologically, the current research utilizes both manual and automated content analyses; highlighting the potential of incorporating computational methods in communication and journalism research to complement and/or corroborate traditional research methods. In the spirit of open science, the datasets and *Python* code have been made freely accessible for other researchers (see previous footnotes).

Contributions to Journalistic Practice

The increasing prominence of online news has led to an evolution in the norms and practices of the journalistic profession (Hedman and Djerf-Pierre 2013). For news outlets, it is useful to know that, despite the negative perceptions regarding clickbaits (Jung et al. 2022), incorporating sensationalist features in tweets generally translates to a higher number of user interactions; in particular, the number of favorites (i.e., likes).

More specifically using emojis, hashtags and media features unequivocally increases the number of likes, based on both the manual and automated content analyses. To encourage higher number of retweets, using slang, hashtags, informal punctuation, all-caps and media in tweets was found to be effective. Conversely, using listicles and forward referencing in tweets negatively impacts both likes and retweets. These findings align with previous research, which found some characteristics of clickbait to be more successful than others in generating engagement (Jung et al. 2022). Also in line with previous research, our findings confirm that entertainment-oriented topics are generally treated more sensationally than politics-oriented topics (see Figure D1 in Online Appendix). The most sensationalist tweets in our manually annotated sample were about sports and entertainment/celebrity news, and the least sensationalist about economy/business.

Conclusion, Limitations, and Future Research

While this study provides useful insights pertaining to the use of sensationalism in news on social media, it has certain limitations. Although *Twitter/X* has been considered the most salient news-oriented platform, future studies can benefit from conducting a comparative analysis across social media platforms. Moreover, a longitudinal analysis, on a larger sample, may also provide useful insights revealing changes in the behavior of digital users over time (i.e., journalists and news audiences). Future studies can also consider other engagement metrics, such as comments or the reactions functionality provided by some platforms, which allow studying emotional responses of users (Lamot, Cools, and Gevers 2024).

Two sensationalist features were difficult to detect (low recall) through automated content analysis: hyperbole and slang. This might be due to the context-specific (e.g., "QB" = quarterback, "Ds" = Democrats) and topic-centric (politics, sports/entertainment) nature of data being analyzed for which the *Python* libraries were not specifically developed for. Scholars have proposed collaboration between computational and social sciences in developing classifiers suited for detecting abstract concepts like clickbait (Molina et al. 2021). Furthermore, the ten features used to detect sensationalism in this research are not exhaustive. Future research can incorporate additional features or dive deeper into individual features. For example, the media feature can be explored in more depth considering the advancement in computer vision technology.

Another limitation is that the current study does not consider the click-through rate: i.e., whether using sensationalist features translates to clicks on the accompanying news items. Researchers could try to incorporate this aspect as well in future analyses. For now, our research demonstrates that sensationalism is widely present in the posts of journalistic organizations on social media, especially the online-native outlets, and that this positively influences user engagement with these posts.

Notes

1. Jupyter notebook with data collection steps, through the *Twitter* API, can be accessed at https://bit.ly/data-extraction-API.
2. Jupyter notebook with Python code for data preprocessing, cleaning and sampling can be accessed at https://bit.ly/3vcBpAa.
3. For codebook, its details and examples, see https://osf.io/djzrx/?view_only=b478b7c6aafa4fa494853ed24dd8c1b8.
4. The results of the first twenty coded tweets were compared and points of confusion were discussed, after which the rest of the sample was coded. Raw data for ICR coding results can be accessed at https://bit.ly/Inter-coder-reliability-data.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Dutch Research Council (NWO) with a Vidi grant under project number: VI.Vidi.211.101 awarded to Dr Mark Boukes.

ORCID

Salman Khawar http://orcid.org/0000-0002-7303-7774
Mark Boukes http://orcid.org/0000-0002-3377-6281

Data Availability Statement

All data used in this research can be openly accessed at dx.doi.org/10.6084/m9.figshare.22132967.

References

Adams, W. C. 1978. "Local Public Affairs Content of TV News." *Journalism Quarterly* 55 (4): 690–695. https://doi.org/10.1177/107769907805500405

Arbaoui, B., K. De Swert, and W. van der Brug. 2020. "Sensationalism in News Coverage: A Comparative Study in 14 Television Systems." *Communication Research* 47 (2): 299–320. https://doi.org/10.1177/0093650216663364

Armstrong, C. L., & Fangfang Gao. (2010). "Now Tweet This: How News Organizations Use Twitter". *Electronic News* 4 (4): 218–235. https://doi.org/10.1177/1931243110389457

Arrese, N., and J. Kaufmann. 2016. "Legacy and Native News Brands Online: Do They Show Different News Consumption Patterns?" *International Journal on Media Management* 18 (2): 75–97. https://doi.org/10.1080/14241277.2016.1200581

Baden, C., C. Pipal, M. Schoonvelde, and M. A. van der Velden. 2021. "Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda." *Communication Methods and Measures* 16 (1): 1–18. https://doi.org/10.1080/19312458.2021.2015574

Berman, D. E. 2021. "The Spaces of Sensationalism: A Comparative Case Study of the New York Journal and BuzzFeed." *International Journal of Communication* 15: 4109–4128. https://ijoc.org/index.php/ijoc/article/view/17352

Beyersmann, E., S. Wegener, and N. Kemp. 2023. "That's Good News ☹: Semantic Congruency Effects in Emoji Processing." *Journal of Media Psychology* 35 (1): 17–27. https://doi.org/10.1027/1864-1105/a000342

Bird, S. 2009. "Tabloidization: What is It, and Does It Really Matter?" In The Changing Faces of Journalism: Tabloidization, Technology and Truthiness, vol. 1, pp. 40–50. London, UK: Routledge. https://doi.org/10.13140/2.1.2223.0404

Biyani, P., K. Tsioutsiouliklis, and J. Blackmer. 2016. "8 Amazing Secrets for Getting More Clicks: Detecting Clickbaits in News Streams Using Article Informality." *Proceedings of the AAAI Conference on Artificial Intelligence* 30 (1): 94–100. https://doi.org/10.1609/aaai.v30i1.9966

Blom, J. N., and K. R. Hansen. 2015. "Click Bait: Forward-Reference as Lure in Online News Headlines." *Journal of Pragmatics* 76: 87–100. https://doi.org/10.1016/j.pragma.2014

Boukes, M., and R. Vliegenthart. 2017. "A General Pattern in the Construction of Economic Newsworthiness? analyzing News Factors in Popular, Quality, Regional, and Financial Newspapers." *Journalism* 21 (2): 279–300. https://doi.org/10.1177/1464884917725989

Boukes, Mark, Bob Van De Velde, Theo Araujo, and Rens Vliegenthart. 2020. "What's the Tone? Easy Doesn't Do It: Analyzing Performance and Agreement between off-THE-Shelf Sentiment Analysis Tools." *Communication Methods and Measures* 14 (2): 83–104. https://doi.org/10.1080/19312458.2019.1671966.

Boumans, J. W., and D. Trilling. 2016. "Taking Stock of the Toolkit: An Overview of Relevant Automated Content Analysis Approaches and Techniques for Digital Journalism Scholars." *Digital Journalism* 4 (1): 8–23. https://doi.org/10.1080/21670811.2015.1096598

Brown, D. K., and V. Sinta. 2016. "Six Things You Didn't Know about Headline Writing: Sensational Form in Viral News of Traditional and Digitally Native News Organizations." *Research Journal of the International Symposium of Journalism*. https://isojjournal.wordpress.com/2016/04/14/six-things-you-didnt-know-about-headlinewriting-sensational-form-in-viral-news-of-traditional-and-digitally-native-news-organizations/.

Brown, D. K., S. Harlow, V. García-Perdomo, and R. Salaverría. 2018. "A New Sensation? an International Exploration of Sensationalism and Social Media Recommendations in Online News Publications." *Journalism* 19 (11): 1497–1516. https://doi.org/10.1177/1464884916683549

Brown, O., L. G. Smith, B. I. Davidson, and D. A. Ellis. 2022. "The Problem with the Internet: An Affordance-Based Approach for Psychological Research on Networked Technologies." *Acta Psychologica* 228: 103650. https://doi.org/10.1016/j.actpsy.2022.103650

Burgers, C., and A. de Graaf. 2013. "Language Intensity as a Sensationalistic News Feature: The Influence of Style on Sensationalism Perceptions and Effects." *Communications – The European Journal of Communication Research* 38 (2): 167–188. https://doi.org/10.1515/commun-2013-0010

Canter, L. 2018. "It's Not All Cat Videos." *Digital Journalism* 6 (8): 1101–1112. https://doi.org/10.1080/21670811.2018.1503058

Chakraborty, A., B. Paranjape, S. Kakarla, and N. Ganguly. 2016. "Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media." 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). https://doi.org/10.1109/ASONAM.2016.7752207

Chakraborty, A., R. Sarkar, A. Mrigen, and N. Ganguly. 2017. "Tabloids in the Era of Social Media? Understanding the Production and Consumption of Clickbaits in Twitter." *SSRN Electronic Journal* 30 (1): 1–21. https://doi.org/10.2139/ssrn.3034591

Chen, Y., N. J. Conroy, and V. L. Rubin. 2015. "Misleading Online Content: Recognizing Clickbait as 'False News'." *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. https://doi.org/10.1145/2823465.2823467

Cowley, S. W. 2017. "The BuzzFeed Marketing Challenge: An Integrative Social Media Experience." *Marketing Education Review* 27 (2): 109–114. https://doi.org/10.1080/10528008.2017.1309981

Curran, J., I. Salovaara-Moring, S. Coen, and S. Iyengar. 2010. "Crime, Foreigners and Hard News: A Cross-National Comparison of Reporting and Public Perception." *Journalism* 11 (1): 3–19. https://doi.org/10.1177/1464884909350640

Firdaus, S. N., C. Ding, and A. Sadeghian. 2021. "Retweet Prediction Based on Topic, Emotion and Personality." *Online Social Networks and Media* 25: 100165. https://doi.org/10.1016/j.osnem.2021.100165

Friedman, B. 2005. "The Penny Press: The Origins of the Modern News Media, 1833–1861." *Journalism History* 31 (1): 147–148. https://doi.org/10.1080/08821127.2005.10677648

García-Perdomo, V., R. Salaverría, D. K. Brown, and S. Harlow. 2017. "To Share or Not to Share." *Journalism Studies* 19 (8): 1180–1201. https://doi.org/10.1080/1461670X.2016.1265896

Gianotto, A. 2016. "Downworthy: A Browser Plugin to Turn Hyperbolic Viral Headlines into What They Really Mean." *downworthy.snipe.net*. Corpus. Retrieved from https://github.com/snipe/downworthy/blob/master/Source/dictionaries/original.js

Hedman, U., and M. Djerf-Pierre. 2013. "The Social Journalist: Embracing the Social Media Life or Creating a New Digital Divide?" *Digital Journalism* 1 (3): 368–385. https://doi.org/10.1080/21670811.2013.776804

Hermida, A., and C. Mellado. 2020. "Dimensions of Social Media Logics: Mapping Forms of Journalistic Norms and Practices on Twitter and Instagram." *Digital Journalism* 8 (7): 864–884. https://doi.org/10.1080/21670811.2020.1805779

Ju, A., S. H. Jeong, and H. I. Chyi. 2013. "Will Social Media save Newspapers?" *Journalism Practice* 8 (1): 1–17. https://doi.org/10.1080/17512786.2013.794022

Jung, A.-K., S. Stieglitz, T. Kissmer, M. Mirbabaie, and T. Kroll. 2022. "Click Me…! the Influence of Clickbait on User Engagement in Social Media and the Role of Digital Nudging." *PLoS One* 17 (6): E0266743. https://doi.org/10.1371/journal.pone.0266743

Kemm, R. 2022. "The Linguistic and Typological Features of Clickbait in YouTube Video Titles." *Social Communication* 8 (1): 66–80. https://doi.org/10.2478/sc-2022-0007

Klein, M., M. Magin, A. A. Riedl, L. Udris, and B. Stark. 2023. "From News Softening to Social News Softening: Comparing Patterns of Political News Coverage on Different (Social) Media Channels in Germany and Switzerland." *Digital Journalism* 1–23. Advance online publication. https://doi.org/10.1080/21670811.2023.2278044

Lamot, K., H. Cools, and I. Gevers. 2024. "Exploring the Relationship between Stylistic Features and Reactions on Facebook: A Comparative Analysis of Newspaper Headlines and Status Messages." *Journalism Studies* 25 (9): 990–1009. https://doi.org/10.1080/1461670X.2024.2359497

Loewenstein, G. 2008. The Psychology of Curiosity: A Review and Reinterpretation. *Psychological Bulletin* 116 (1): 75–98. https://doi.org/10.1037//0033-2909.116.1.75

Mbaye, M. 2019. "Up-to-Date List of Slangs for Text Preprocessing." *Kaggle*. https://www.kaggle.com/code/nmaguette/up-to-date-list-of-slangs-for-text-preprocessing

McCartney, B. 2016. "Reflective Eye." *Editor & Publisher* 149 (10): 16–19.

Metzger, M. J., A. J. Flanagin, and R. B. Medders. 2010. "Social and Heuristic Approaches to Credibility Evaluation Online." *Journal of Communication* 60 (3): 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Molek-Kozakowska, K. 2013. "Towards a Pragma-Linguistic Framework for the Study of Sensationalism in News Headlines." *Discourse & Communication* 7 (2): 173–197. https://doi.org/10.1177/1750481312471668

Molina, M. D., S. S. Sundar, M. M. Rony, N. Hassan, T. Le, and D. Lee. 2021. "Does Clickbait Actually Attract More Clicks? Three Clickbait Studies You Must Read." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411764.3445753

Murtha, J. 2015. "What It's like to Get Paid for Clicks." *Columbia Journalism Review*. https://www.cjr.org/analysis/the_mission_sounds_simple_pay.php.

Neilson, T. 2012. "Journalists Strike Online: Visibility, Field and the Huffington Post." *Global Media Journal* 12 (20): 1–10. https://www.proquest.com/scholarly-journals/journalists-strike-online-visibility-field/docview/1323984178/se-2?accountid=14615

Newman, N., R. Fletcher, K. Eddy, C. T. Robertson, and R. K. Nielsen. 2023. (rep.). *Reuters Institute Digital News Report 2023* (p. 109).

Novak, P., J. Smailović, B. Sluban, and I. Mozetič. 2015. "Sentiment of Emojis." *PLoS One* 10 (12): E0144296. https://doi.org/10.1371/journal.pone.0144296

Orellana-Rodriguez, C., D. Greene, and M. T. Keane. 2017. "Spreading One's Tweets: How Can Journalists Gain Attention for Their Tweeted News?" *Journal of Web Science* 3 (1): 16–31. https://doi.org/10.1561/106.00000009

Otto, L., I. Glogger, and M. Boukes. 2016. "The Softening of Journalistic Political Communication: A Comprehensive Framework Model of Sensationalism, Soft News, Infotainment, and Tabloidization." *Communication Theory* 27 (2): 136–155. https://doi.org/10.1111/comt.12102

Pelau, C., M.-I. Pop, M. Stanescu, and G. Sanda. 2023. "The Breaking News Effect and Its Impact on the Credibility and Trust in Information Posted on Social Media." *Electronics* 12 (2): 423. https://doi.org/10.3390/electronics12020423

Petty, R. E., and J. T. Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. In Communication and Persuasion (Ser. Springer Series in Social Psychology) pp. 1–24. New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4964-1_1

Potthast, M., S. Köpsel, B. Stein, and M. Hagen. 2016. "Clickbait Detection." Advances in Information Retrieval – *ECIR* 2016 (9626): 810–817. https://doi.org/10.1007/978-3-319-30671-1_72

Potthast, M., T. Gollub, K. Komlossy, S. Schuster, M. Wiegmann, E. Fernandez, M. Hagen, and B. Stein. 2018. "Crowdsourcing a Large Corpus of Clickbait on Twitter." Proceedings of the 27th International Conference on Computational Linguistics, 1498–1507.

Puspitorini, Ferawaty. 2020. "An Analysis of Slang Words Used in Twitter." In The Second International Conference on Progressive Education (ICOPE), 17–18 October, Universitas Negeri Lampung.

Riordon, K. 2014. *Accuracy, Independence, and Impartiality: How Legacy Media and Digital Natives Approach Standards in the Digital Age*. Oxford, UK: Oxford University.

Rony, M. M., N. Hassan, and M. Yousuf. 2017. "Diving Deep into Clickbaits." *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. https://doi.org/10.1145/3110025.3110054

Ryfe, D. M. 2009. "Broader and Deeper." *Journalism* 10 (2): 197–216. https://doi.org/10.1177/1464884908100601

Sadri, S. R. 2019. "Listicles and the Modern News Article." *Atlantic Journal of Communication* 27 (2): 83–98. https://doi.org/10.1080/15456870.2019.1574794

Safori, A. O. 2019. "Journalist Use of Social Media: Guidelines for Media Organizations." *The Journal of Social Sciences Research* 54: 1061–1068. https://doi.org/10.32861/jssr.54.1061.1068

Scacco, J. M., and A. Muddiman. 2019. "The Curiosity Effect: Information Seeking in the Contemporary News Environment." *New Media & Society* 22 (3): 429–448. https://doi.org/10.1177/1461444819863408

Scott, K. 2021. "You Won't Believe What's in This Paper! Clickbait, Relevance and the Curiosity Gap." *Journal of Pragmatics* 175: 53–66. https://doi.org/10.1016/j.pragma.2020.12.023

Shi, J., K. K. Lai, and G. Chen. 2023. "Examining Retweeting Behavior on Social Networking Sites from the Perspective of Self-Presentation." *PLoS One* 18 (5): e0286135. https://doi.org/10.1371/journal.pone.0286135

Siegert, G., K. Förster, S. M. Chan-Olmsted, and M. Ots. 2016. *Handbook of Media Branding* (Softcover reprint of the original 1st ed., 2015 ed.). Berlin: Springer.

Skovsgaard, M. 2014. "A Tabloid Mind? Professional Values and Organizational Pressures as Explanations of Tabloid Journalism." *Media, Culture & Society* 36 (2): 200–218. https://doi.org/10.1177/0163443713515740

Song, H., P. Tolochko, J.-M. Eberl, O. Eisele, E. Greussing, T. Heidenreich, F. Lind, S. Galyga, and H. G. Boomgaarden. 2020. "In Validations we Trust?" *Political Communication* 37 (4): 550–572. https://doi.org/10.1080/10584609.2020.1723752

Staender, A., E. Humprecht, F. Esser, S. Morosoli, and P. Van Aelst. 2021. "Is Sensationalist Disinformation More Effective? Three Facilitating Factors at the National, Individual, and Situational Level." *Digital Journalism* 10 (6): 976–996. https://doi.org/10.1080/21670811.2021.1966315

Stanton, R. (Ed.). 2009. "Editor Announces Weekly Tabloid Aimed at Social-Networking Readers". *Los Angeles Times*, March 25. https://latimesblogs.latimes.com/readers/2009/03/editor-announce.html

Subramanian, K. 2017. "Product Promotion in an Era of Shrinking Attention Span." *International Journal of Engineering and Management Research* 7: 85–91.

Sui, M., I. Hawkins, and R. Wang. 2023. "When Falsehood Wins? Varied Effects of Sensational Elements on Users' Engagement with Real and Fake Posts." *Computers in Human Behavior* 142: 107654. https://doi.org/10.1016/j.chb.2023.107654

Uribe, R., and B. Gunter. 2007. "Are 'Sensational' News Stories More Likely to Trigger Viewers' Emotions than Non-Sensational News Stories?" *European Journal of Communication* 22 (2): 207–228. https://doi.org/10.1177/0267323107076770

Van Dijck, J., and T. Poell. 2013. "Understanding Social Media Logic." *Media and Communication* 1: 1. https://doi.org/10.12924/mac2013.01010002

Van Duyn, E., and J. Collier. 2018. "Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media." *Mass Communication and Society* 22 (1): 29–48. https://doi.org/10.1080/15205436.2018.1511807

Vosoughi, S., D. Roy, and S. Aral. 2018. "The Spread of True and False News Online." *Science (New York, N.Y.)* 359 (6380): 1146–1151. https://doi.org/10.1126/science.aap9559

Walker, M., and K. E. Matsa. 2021. *News Consumption Across Social Media in 2021*. Pew Research Center's Journalism Project. September 20. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/

Walter, T. 2019. "Twitter Slang." *Kaggle*. https://www.kaggle.com/datasets/gogylogy/twitterslang

Watson, A. 2023. "Leading Print Newspapers U.S. by Circulation 2023." *Statista*. August 22. https://www.statista.com/statistics/272790/circulation-of-the-biggest-daily-newspapers-in-the-us/

Welbers, K., and M. Opgenhaffen. 2018. "Presenting News on Social Media." *Digital Journalism* 7 (1): 45–62. https://doi.org/10.1080/21670811.2018.1493939