How are negative articles consumed? A quantitative analysis of user behavior in a real news service

Kazuya Ohata*, Hitoshi Iyatomi*, Hajime Morita†, Kojiro Iizuka†
* *Department of Applied Informatics, Graduate School of Science and Engineering, Hosei Univeristy*, Tokyo, Japan
kazuya.ohata.2b@gmail.com, iyatomi@hosei.ac.jp
† *Gunosy Inc.*, Tokyo, Japan
{hajime.morita, kojiro.iizuka}@gunosy.com

***Abstract*—Recommendation algorithms automatically suggest news articles based on past behavioral logs. There has been reported cases of mental health problems caused by continuous consumption of negative articles, besides recommendation algorithms has a problem of over-recommendation which may induce continuous consumption of negative articles. Although research on the relationship between negative news article consumption and mental health has been conducted via small-scale user interviews, large-scale behavioral research on user engagement with recommended news articles has not been carried out. Therefore, we comprehensively investigated how the emotional polarity of articles affects each indicator of user attention by assigning emotional labels to news articles using crowdsourcing and analyzing about 1 million user behavior logs that viewed these articles. To the best of our knowledge, this is one of the first publicly available studies to analyze the impact of negative articles on users' news consumption behavior on an online news platform. The findings indicated that negative articles, irrespective of their category, were more likely to be clicked on, were read for longer durations, and had lower bounce rates. Furthermore, users showed greater interest in negative news related to entertainment and sports. These findings can be used as a first step for news platforms to build safer recommendation algorithms that consider the psychological impact on users.**

***Index Terms*—news services, sentiment analysis, doomscrolling**

I. Introduction

News recommendation services are popular because they provide personalized news content that users find interesting. However, they also pose the risk of negative health effects on users. Personalized algorithmic recommendations can limit the range of information users are exposed to by over-recommending high-profile, similar, and stimulating articles. This can narrow their thinking and sway complex human emotions, negatively impacting their mental and physical health. Recent studies have reported that users who consume more negative news articles are more likely to experience mental and physical ill health [1].

Although there have been improvements in recommendation algorithms in the news distribution field that take into account emotional polarity, to the best of our knowledge, no large-scale user behavior analysis has been done using labels that reflect human perception. Several recommendation systems have been proposed that consider the emotional diversity of users, but, to our understanding, no academic reports have utilized that actual service data in a recommendation system that includes the diversity of user emotions [2], [3]. Some analysis have been performed on doomscrolling and health status, which involves searching for and consuming negative news articles, but these have been limited to small-scale user interviews aimed at providing information in the medical field [4]. To obtain a more comprehensive understanding of user behavior, we think it is essential to conduct a more in-depth analysis not only of whether users are interested in news articles but also of articles that demonstrate sustained interest, as indicated by dwell time, and articles that lose interest immediately. By conducting such an analysis, we can gain insights into the specific factors that drive user behavior and use this information to improve news recommendation algorithms and mitigate the negative health effects of consuming news.

In this paper, we employed a reliable crowdsourcing platform to assign an emotional polarity (positive, neutral, or negative) to 5,000 articles on a commercial news platform with approximately 1 million users and conducted a comprehensive analysis of the impact of negative articles on user behavior using over a month's worth of data. This paper is one of the first research publications to analyze the impact of negative articles on actual user behavior. To investigate the possibility of bias in the attention to and consumption of negative information, we analyzed three representative indicators of user behavior commonly used in the analysis of recommendation systems [5], [6]: (1) click-through rate (CTR), (2) dwell time, and (3) bounce rate. Specifically, this paper conducted an extensive survey and discussion of the following research questions.

- RQ1. What differences in CTR, dwell time, and bounce rate are observed for different emotional polarities?
- RQ2. How does the relationship between time spent and the number of words differ by emotional polarity?

The contribution of this paper is the analysis of a large user behavior log on the relationship between emotion and the consumption of news articles, utilizing data with emotional polarity labels determined by human cognition.

II. Related Work

A. Negative health effects of consuming negative information

The association between the continuous consumption of strongly polarizing articles and adverse health effects has been actively studied in recent years [7]–[10]. Studies conducted during the COVID-19 pandemic reported that viewing articles about COVID-19 through websites and social media can cause anxiety, depression, post-traumatic stress disorder (PTSD), and other psychiatric disorders [11]. In addition to COVID-19, exposure to articles about war, conflict, and disasters has been reported to be associated with increased early symptoms of PTSD. Ytre-Arne and Moe conducted a survey and found that overexposure to negative article topics not only drains the psyche but also accelerates the doomscroll phenomenon of fishing for more negative articles [12]. Furthermore, extensive coverage of celebrity suicides has been found to increase suicide rates, prompting the media to review the frequency and content of their coverage [13], [14]. While it is undeniable that there are differences in reporting trends among countries and regions, we believe that, in today's world of instantaneous information dissemination, the regional specificity of news articles is becoming smaller, and the problems caused by negative articles are common to all regions.

B. Negative impacts of biased information

News with a highly emotional polarity bias often includes the reporter's emotions and personal opinions. Presenting multiple such news stories may provide readers with biased perspectives and stimuli. Regina et al. studied young people's consumption of news and found that they preferred biased news of their own taste rather than news of a public nature [15]. They tended to prefer news that included the author's opinions and emotions rather than news articles with an objective point of view. Strongly opinionated articles can narrow users' perspectives, make it difficult for them to consider different points of view, and cause a division between the community and the public [16], [17]. Information eclipse can also cause the so-called filter bubble problem in which algorithms biasedly recommend similar articles according to past user behavior logs, thus accelerating the recommendation of highly stimulating articles and leading to a cycle of providing biased viewpoints. Tomlein et al. reported that, in response to the filter bubble problem in misinformation on YouTube, one can escape from this problem by viewing videos that negate the misinformation [18]. Getting out of this information vortex is not an easy task, as it requires users to choose the information to which they are exposed to, and this may require an approach from service providers. In the case of news programs and newspapers, where the content that users are exposed to is managed by humans, it is possible to strike a balance by placing positive information next to negative information. However, online news providers that deliver news mechanically, without human intervention, are not operating with this point in mind. In order to prevent such information bias, diversity has recently been evaluated [19], [20], and a content recommendation method has been proposed to prevent bias in polarity when diversifying emotional polarity [21]. On the other hand, these efforts targeting the emotional polarity of news articles have been limited to simulations using mechanically generated simulated data or user interviews with small sample sizes in surveys targeting actual users. In contrast to these previous studies, we investigated user behavior across multiple metrics on a large news platform using article labels based on human cognition.

III. Method

A. Target news service

In this study, we analyzed Gunosy[1], one of Japan's most well-known news recommendation applications. In this application, users can browse a list of article titles on the home page and freely click on articles of interest to view their full text. The user can return to the home page at any time while browsing the content of an article and can search for the article of interest again. The order of the articles on the home page is such that the top 1 to 5 articles are topical and public articles selected by the editors, while the rest are recommended by a personalization algorithm that differs for each individual. This study aimed to analyze the consumption of articles recommended by the algorithm, and articles selected by editors were excluded from the analysis.

B. Dataset

We collected data on approximately 300,000 articles over the course of one month on the application. The news articles posted by the site come from over 100 media outlets and sources, including traditional media outlets such as television stations and newspapers, as well as emerging Internet news sites. While news articles may capture readers' attention with exaggerated titles, they may also contain content that differs from the title or contain clickbait that is not substantial in the body of the article [22]. However, as the service we used in this study includes only articles from reputable publishers that have undergone content screening, there is a certain degree of guarantee of correspondence between the article title and the article text content. We used the top 1,000 articles viewed per day to create a dataset of about 120,000 articles and user behavior logs between 30 June and 31 July 2021. From these, we randomly selected 1,250 articles each from society, entertainment, sports, and science and technology (ScienceTech) categories to create a dataset of 5,000 articles.

C. Labels for emotional polarity

As depicted in Fig. 1, we utilized Lancers[2], a major crowdsourcing site in Japan, to gather the labeling outcomes of five workers for each of the 5,000 articles. Each article was assigned one of three emotional polarity labels (positive, neutral, or negative), following the labeling rules established

[1] https://gunosy.com/

[2] https://www.lancers.jp/

Rule

❏ Select the "**negative**" label for articles that are likely to cause **anxiety, shock, or gloom** from the headline, such as the following,

- Socially unsettling events, such as disasters, casualties from accidents, infectious disease risks, and violent or criminal acts.
  - Man dies after falling at the construction site.
  - Coronavirus in Tokyo records the highest number of infections this month.
- The scandal caused by a famous person, report of inappropriate behavior, exposure of private life such as separation, news of a death, criticism of appearance.
  - Mr. X is in big trouble with a woman he cheated on.
  - Actor Y died at the age of 83.
- Figures, analysis, and commentary on politics, economics, and international issues.
  - The bitcoin crash caused chaos around the world.
  - Ukraine's military command declines, fighting may become more protracted.
- Loss of a game or injury of a player in the sports field.
  - Giants failed to stop their losing streak.
  - Player X had to withdraw from a tournament due to a right leg injury.
- Problems and incidents related to science and technology.
  - Cyber attack on a multinational beverage company, affecting production.
  - Leak of 100,000 customers' data at Y company.

❏ Select the "**positive**" label for articles expected to address a positive matter that receives **reassurance, cheerfulness, or brightness** from the headline.

- Socially reassuring events, such as rescue, restoration of condition, or reduction of infection in a disaster or incident.
  - Emergency crews rescue five stranded people.
  - Coronavirus infection decreased for the 10th consecutive day.
- Record breaking, the release of works, assumption of a role, life events such as marriage or pregnancy, or appearance accolades regarding a celebrity.
  - X releases a new music video today.
  - Mr. X announces his marriage! The partner is a famous actor, Y.
  - Former idol X undergoes a dramatic change, becoming mature and calm.
- Figures and analysis on political, economic, and international issues where the outlook is favorable.
  - Major corporations increased summer bonuses by an average of 8%.
  - Change of government expectations for better relations with neighboring countries.
- Positive articles in the field of sports, such as game victories and athletes' successes.
  - Gold medal for women's softball at the Tokyo Olympics.
- Innovations in science and technology.
  - Company X is further streamlining its operations by introducing image recognition AI.
  - iPhones with the latest chips will go on sale in October.
- Leisure spots, gourmet foods, and products.
  - The best places to go this summer! Top 5 Sightseeing Spots.
  - Introduction of restaurants with good beer.

❏ Select the "**neutral**" label for articles that do not appear to fall into either of the above categories.

- A deep-sea fish, the oarfish, was captured.
- The number one job high school students want to be is a public servant.
- Australia's national soccer team plays to a 0-0 draw.

Fig. 1. Description shown to workers on the crowdsourcing task.

TABLE I
EXAMPLES OF ARTICLE TITLES AND INFERENCE POLARITY LABELS

Title	Label
X and Y celebrate 80 years of marriage.	Positive
Large whales appeared on X beach yesterday.	Neutral
Scenes from a destroyed school: 20 people killed by dropped bombs.	Negative
Was X in poor health? On the sudden death of a famous actor.	Negative
What caused the chaos that erupted in the Champions League?	Negative

by our researchers. When we recruited workers to obtain labels that more closely resembled the actual users, we attracted nearly 100 men and women in their 20s to 60s from various occupations. After completion, we aggregated the labels and found that 3,358 out of 5,000 articles (approximately 67%) received the same label from at least four of the five raters. Few labels were divided between positive and negative by multiple raters. This high concordance rate indicates that the sentiments received from news articles are generally consistent regardless of social status, while reflecting, to some extent, the diversity of individual workers in different social settings. To aggregate labels and obtain emotion labels for each article, we used the generative model of labels, abilities, and difficulties (GLAD) [23], which uses the expectation–maximization (EM) algorithm [24] to estimate the difficulty of the task and the ability parameters of the worker and predicts the true correct answer from these values. Table I shows examples of articles used in this study.

Table II shows the number of articles per category. Although the articles were randomly sampled, it was observed that there were more articles with negative labels in the society category and more articles with positive labels in the ScienceTech category.

IV. Analysis

For these 5,000 articles, we conducted an analysis to answer the following two research questions with the following evaluation indicators:

- RQ1. What differences in CTR, dwell time, and bounce rate are observed for different emotional polarities?

TABLE II
DATA BREAKDOWN AFTER AGGREGATING LABELS FROM THE CROWDSOURCING RESULTS

Categories	Positive	Neutral	Negative
Society	254	273	723
Entertainment	690	299	261
Sports	739	194	317
ScienceTech	891	272	87
Total	2,574	1,038	1,388

- RQ2. How does the relationship between time spent and the number of words differ by emotional polarity?

A. Indicators

- **CTR:** CTR is the ratio of the number of clicks to the number of times content is viewed and is a common metric used to optimize content placement and search rankings. A high CTR indicates that many users find the content relevant and interesting, and that personalized articles match user preferences.
- **Dwell Time:** Dwell time represents the time from when a user clicks on an article title, accesses a web page or content, and leaves, and it is one of the common indicators for understanding user satisfaction. A longer dwell time indicates that the web content is providing valuable and engaging information to the user.
- **Bounce Rate:** The bounce rate is the percentage of visitors who leave the article page shortly after accessing it. A high bounce rate indicates that users quickly lose interest in the article. Considering that the news articles used in this experiment have a smaller volume of content than typical web content, and the distribution of total time users spend in the article body, we set the threshold to 2.0 seconds.
- **Article Length:** This indicator represents the length of the article body (number of characters) and is intended to investigate the relationship with dwell time. Since the number of headlines, images, and layout of published articles are typically standardized in the news services used in this study, the dwell time is expected to be less affected by factors other than article length.

B. Analytical Method

To avoid the business risk of disclosing the raw indicators, we normalized the data by setting their mean value to 1 and showed their relative differences.

1) RQ1: Differences in index values by emotion: We examined the differences in trends between categories for the three emotional polarities of positive/neutral/negative and identified which polarity pairs showed differences. As the distribution of each indicator for all articles did not follow a normal distribution, the Mann–Whitney U test [25] was used instead of the t-test. Since we conducted pairwise tests for each category, a total of 12 tests were carried out, and the Benjamini–Hochberg method [26] was applied to adjust the p-values of each pair, with a false discovery rate of 0.05 for multiple testing correction. It also presents effect sizes calculated by cliff delta [27], using the U statistics, with absolute values of 0.147 indicating a small effect size, 0.33 indicating a medium effect size, and 0.474 indicating a large effect size, in comparison to the commonly used Cohen's d effect sizes [28] of 0.2 (small), 0.5 (medium), and 0.8 (large) [29]. In this experiment, based on the above, pairs showing an effect size of approximately 0.3 or greater were highlighted as a rough guide to the effect size.

2) RQ2: Emotions and their relationship to dwell time and the article length: Based on the hypothesis that longer articles tend to have longer dwell times, we examined the relationship between the number of words in the body of an article and dwell time in relation to emotional polarity.

V. RESULTS

A. RQ1: Differences in index values by emotion

Fig. 2 shows box-and-whisker plots for evaluation indicators by news category and emotional polarity, with the upper and lower bounds ($1.5\times$IQR[interquartile range]) away from each quartile point. Regarding the CTR (Fig. 2a), the CTR of negative articles was more highly distributed than that of positive and neutral articles, and the difference in means was statistically significant at $p < 0.001$ for both positive and neutral articles. In terms of dwell time (Fig. 2b), negative articles generally had longer dwell times, except for those in the society category. In all news categories, pairs with different emotional polarities showed statistically significant differences. As for the bounce rate (Fig. 2c), there was a decreasing trend in the IQR range as the emotional polarity of the articles became more negative in all news categories. The statistical significance of the differences in means varied by category. Table III summarizes the p-values and the effect sizes by Cliff's delta. Many of the pairs were statistically significant when the effect size was large.

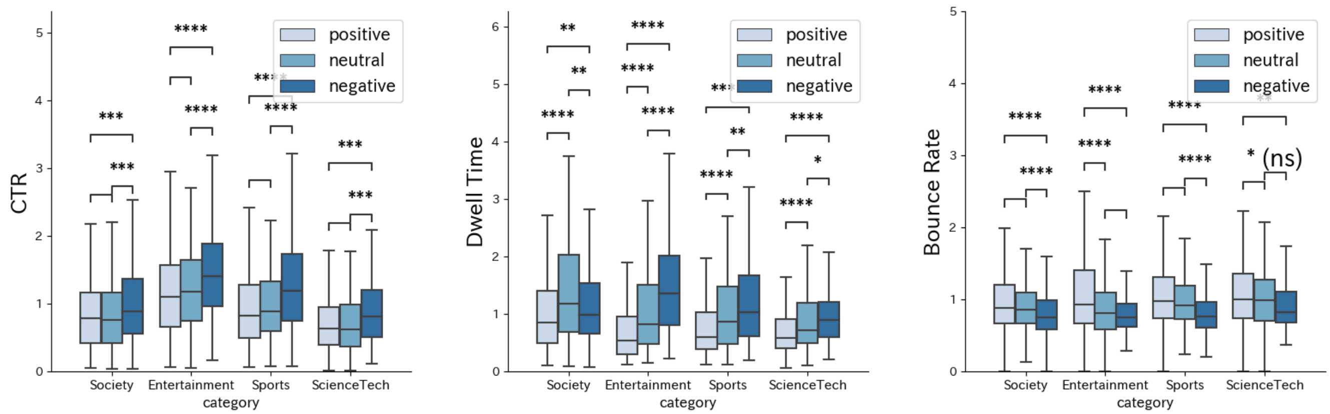
B. RQ2: Emotions and their relationship to dwell time and the article length

Table IV summarizes the number of characters in the body of the article and Pearson's correlation coefficients between word count and dwell time.

VI. DISCUSSION

A. Differences in index values by emotion

In all categories, negative articles had a higher CTR than articles with other polarities, indicating that users are more likely to click on negative articles when they are displayed in the list of articles. In addition, in many categories, the dwell time of negative articles is greater than that of articles with other polarities, while the bounce rate is lower for negative articles, indicating that users stay in the articles longer. This suggests that users are more likely to engage with negative articles and to stay on the page longer after clicking on the title of a negative article. As shown in Table III, the areas where the effect size stands out are mainly in the entertainment



(a) CTR for each category. Negative labels for all categories are significantly higher than other labels.

(b) Dwell time for each category. Significant differences are found in all categories for all polarity pairs.

(c) Bounce rate for each category. In all categories, values decrease as label polarity goes from positive to negative, but pairs with significant differences vary by category.

Fig. 2. User activities for each emotional polarity label. We conducted Mann-Whitney U test with the Benjamini–Hochberg correction. In each figure, asterisks denote p-values: * for $0.01 \leq p < 0.05$, ** for $0.001 \leq p < 0.01$, *** for $0.0001 \leq p < 0.001$, and **** for $p < 0.0001$. Note that the values are displayed as multiples of the mean value of 1.0 since there is a business risk to exposing raw metrics.

TABLE III
DIFFERENCES IN MEANS AMONG DIFFERENT POLAR PAIRS WITHIN EACH INDICATOR

Category	Group Pair	CTR		Dwell Time		Bounce Rate	
		p-value	†effect size	p-value	†effect size	p-value	†effect size
Society	Pos. vs. Neg.	3.940E-04 ***	−0.165	1.533E-03 **	−0.149	1.119E-06 ****	0.206
	Pos. vs. Neu.	9.940E-01	−0.018	5.759E-06 ****	−0.244	5.175E-01	0.033
	Neu. vs. Neg.	2.791E-04 ***	−0.167	4.897E-03 **	0.093	5.866E-06 ****	0.186
Entertainment	Pos. vs. Neg.	4.051E-08 ****	−0.230	8.002E-42 ****	**−0.569**	4.574E-11 ****	0.276
	Pos. vs. Neu.	2.870E-01	−0.049	1.831E-14 ****	**−0.312**	8.447E-07 ****	0.197
	Neu. vs. Neg.	6.593E-05 ****	−0.201	3.943E-09 ****	**−0.293**	1.592E-01	0.069
Sports	Pos. vs. Neg.	1.160E-13 ****	**−0.314**	2.038E-24 ****	**−0.419**	1.244E-16 ****	**0.321**
	Pos. vs. Neu.	2.591E-01	−0.084	5.610E-07 **	−0.260	1.237E-01	0.072
	Neu. vs. Neg.	2.242E-06 ****	−0.274	7.121E-03 ****	−0.169	2.307E-07 ****	0.273
ScienceTech	Pos. vs. Neg.	7.522E-04 ***	−0.222	9.220E-08 ****	**−0.350**	2.660E-03 **	0.195
	Pos. vs. Neu.	4.903E-01	0.016	4.028E-05 ****	−0.175	3.029E-01 *	0.037
	Neu. vs. Neg.	7.280E-04 ***	−0.246	2.476E-02 *	−0.166	3.507E-02	0.146

† The effect size with an absolute value of approximately 0.3 or more is shown in bold.

and sports categories, which means that people are particularly enthusiastic about negative articles in the entertainment and sports fields. Regarding RQ1, articles with a negative polarity were more likely to be consumed for a longer period of time and at a higher rate than articles with other polarities, but this trend varied by category.

B. Emotions and their relationship to dwell time and the article's body length

For categories other than ScienceTech, a correlation of 0.4 or higher was observed between the length of the article and dwell time for negative articles, suggesting that negative articles tend to hold users' attention longer as the length of the article increases. However, in the entertainment and sports categories, negative content had a longer dwell time, even when the article length was shorter than that of other polarities, suggesting that negative content efficiently captures users' attention even with a shorter article length. Other than negative articles, other articles with high correlation values include positive articles in the society category and neutral articles in the sports category, which may attract users by making these articles more substantial and with larger content. In response to RQ2, there is an overall positive correlation between article length and time spent, regardless of polarity, with negative articles in many categories showing a slightly stronger correlation of $r = 0.4$ or greater.

VII. CONCLUSION

In this study, we present the first comprehensive analysis of the impact of negative news articles on user behavior using a large dataset of news articles. Our results indicate that articles with negative polarity are more likely to be clicked,

TABLE IV
RELATIONSHIP BETWEEN DWELL TIME AND NUMBER OF CHARACTERS

Category	Society			Entertainment			Sports			ScienceTech		
Sentiment	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
Character Count	0.834	1.441	1.207	0.910	0.991	0.858	0.771	1.018	0.739	1.177	1.463	1.188
Correlation with Dwell Time	0.529	0.309	0.507	0.277	0.169	0.426	0.357	0.559	0.445	0.365	0.268	0.184

stay longer, and have a lower bounce rate than articles with other polarities. In addition, many categories show a positive correlation of over 0.4 between the number of characters and the time spent on negative articles, suggesting that users are more engaged with the content. These findings suggest that the emotional polarity of articles has a significant impact on user behavior and can inform efforts to create a safer platform for both news providers and their users.

REFERENCES

[1] B. McLaughlin, M. R. Gotlieb, and D. J. Mills, "Caught in a dangerous world: Problematic news consumption and its relationship to mental and physical ill-being," *Health Communication*, vol. 0, pp. 1–11, 2022.

[2] C. Wu, F. Wu, T. Qi, and Y. Huang, "SentiRec: Sentiment diversity-aware neural news recommendation," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, Dec. 2020, pp. 44–53.

[3] H. Han, C. Wang, Y. Zhao, M. Shu, W. Wang, and Y. Min, "SSLE: A framework for evaluating the "filter bubble" effect on the news aggregator and recommenders," *World Wide Web*, vol. 25, no. 3, pp. 1169–1195, 2022.

[4] M. Price, A. C. Legrand, Z. M. Brier, K. van Stolk-Cooke, K. Peck, P. S. Dodds, C. M. Danforth, and Z. W. Adams, "Doomscrolling during COVID-19: The negative association between daily social and traditional media consumption and mental health symptoms during the COVID-19 pandemic,," *Psychological Trauma: Theory, Research, Practice, and Policy*, vol. 14, no. 8, 2022.

[5] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proceedings of the 15th International Conference on Intelligent User Interfaces.*, 2010, pp. 31–40.

[6] D. Lagun and M. Lalmas, "Understanding user attention and engagement in online news reading," in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 2016, pp. 113–122.

[7] J. Gao, P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, and J. Dai, "Mental health problems and social media exposure during COVID-19 outbreak," *PLOS One*, vol. 15, no. 4, p. e0231924, 2020.

[8] M. Y. Ni, L. Yang, C. M. Leung, N. Li, X. I. Yao, Y. Wang, G. M. Leung, B. J. Cowling, and Q. Liao, "Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: Cross-sectional survey," *JMIR Mental Health*, vol. 7, no. 5, p. e19009, 2020.

[9] K. Stainback, B. N. Hearne, and M. M. Trieu, "COVID-19 and the 24/7 news cycle: Does COVID-19 news exposure affect mental health?" *Socius*, vol. 6, p. 2378023120969339, 2020.

[10] J. K. Kellerman, J. L. Hamilton, E. A. Selby, and E. M. Kleiman, "The mental health impact of daily news exposure during the COVID-19 pandemic: Ecological momentary assessment study," *JMIR Mental Health*, vol. 9, no. 5, 2022.

[11] Y. Neria and G. M. Sullivan, "Understanding the mental health effects of indirect exposure to mass trauma through the media," *JAMA*, vol. 306, no. 12, pp. 1374–1375, 2011.

[12] B. Ytre-Arne and H. Moe, "Doomscrolling, monitoring and avoiding: News use in COVID-19 pandemic lockdown," *Journalism Studies*, vol. 22, no. 13, pp. 1739–1755, 2021.

[13] A. T. Cheng, K. Hawton, C. T. Lee, and T. H. Chen, "The influence of media reporting of the suicide of a celebrity on suicide rates: A population-based study," *International Journal of Epidemiology*, vol. 36, no. 6, pp. 1229–1234, 2007.

[14] Y. J. Choi and H. Oh, "Does media coverage of a celebrity suicide trigger copycat suicides?: Evidence from korean cases," *Journal of Media Economics*, vol. 29, no. 2, pp. 92–105, 2016.

[15] R. Marchi, "With facebook, blogs, and fake news, teens reject journalistic "objectivity"," *Journal of Communication Inquiry*, vol. 36, no. 3, pp. 246–262, 2012.

[16] S. Flaxman, S. Goel, and J. M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public Opinion Quarterly*, vol. 80, no. S1, pp. 298–320, 2016.

[17] D. Baden, K. McIntyre, and F. Homberg, "The impact of constructive news on affective and behavioural responses," *Journalism Studies*, vol. 20, no. 13, pp. 1940–1959, 2019.

[18] M. Tomlein, B. Pecher, J. Simko, I. Srba, R. Moro, E. Stefancova, M. Kompan, A. Hrckova, J. Podrouzek, and M. Bielikova, "An audit of misinformation filter bubbles on YouTube: Bubble bursting and recent behavior changes," in *Fifteenth ACM Conference on Recommender Systems*, 2021, pp. 1–11.

[19] T. Qi, F. Wu, C. Wu, and Y. Huang, "PP-Rec: News recommendation with personalized user interest and time-aware news popularity," *arXiv preprint arXiv:2106.01300*, 2021.

[20] M. Alam, A. Iana, A. Grote, K. Ludwig, P. Müller, and H. Paulheim, "Towards analyzing the bias of news recommender systems using sentiment and stance detection," in *Companion Proceedings of the Web Conference 2022*, 2022, pp. 448–457.

[21] K. Garimella, A. Gionis, N. Parotsidis, and N. Tatti, "Balancing information exposure in social networks," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[22] V. Kumar, D. Khattar, S. Gairola, Y. Kumar Lal, and V. Varma, "Identifying clickbait: A multi-strategy approach using neural networks," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1225–1228.

[23] J. Whitehill, T.-f. Wu, J. Bergsma, J. Movellan, and P. Ruvolo, "Whose vote should count more: Optimal integration of labels from labelers of unknown expertise," *Advances in Neural Information Processing Systems*, vol. 22, pp. 2035–2043, 2009.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[25] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[26] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[27] N. Cliff, "Dominance statistics: Ordinal analyses to answer ordinal questions." *Psychological Bulletin*, vol. 114, no. 3, p. 494, 1993.

[28] J. Cohen, "Statistical power analysis," *Current Directions in Psychological Science*, vol. 1, no. 3, pp. 98–101, 1992.

[29] J. Romano, J. D. Kromrey, J. Coraggio, J. Skowronek, and L. Devine, "Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and cohen's d indices the most appropriate choices?" in *Annual Meeting of the Southern Association for Institutional Research*, 2006, pp. 1–51.