

Examining the Role of Clickbait Headlines to Engage Readers with Reliable Health-related Information

Naeemul Hassan



UNIVERSITY OF
MARYLAND

IS

J



Md Main Uddin Rony



Naeemul Hassan



Md Mahfuzul Haque



THE UNIVERSITY of
MISSISSIPPI

J



Sima Bhowmik



Kristen A Swain

72% of adult internet users in U.S.
searched online for information
about health issues.

-Pew Research, 2014

<https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/>

Both Reliable and Unreliable
Health news are Competing
for Your attention

More than **50%** of the top-20 Facebook stories containing “cancer” in headline were **False**.

--BuzzFeed.News, 2017

<https://www.buzzfeednews.com/article/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in>

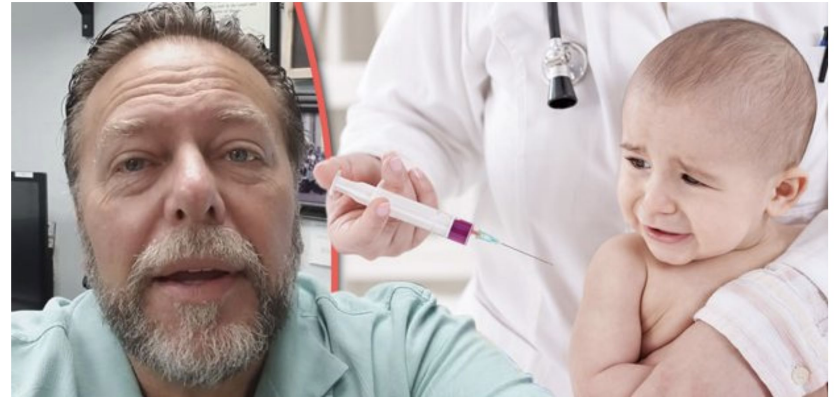
Doctor Blows Whistle on Flu Shot: 'It's Designed to Spread Cancer'

April 26, 2018 by Edward Morgan

f Share

Twitter Tweet

Newsletter Signup



Dr. John Bergman issues warning to the public of 'flu panic'. Dr. John Bergman says the flu

Reliable vs Unreliable: What are the Differences?

Insomnia in adults linked to heart attack and stroke

By Sandee LaMotte, CNN

⌚ Updated 5:17 PM ET, Wed November 6, 2019



Doctor Blows Whistle on Flu Shot: 'It's Designed to Spread Cancer'

April 26, 2018 by Edward Morgan

[f Share](#) [Tweet](#) [Newsletter Signup](#)



Dr. John Bergman issues warning to the public of 'flu panic'. Dr. John Bergman says the flu

11 Health On the Net

8 U.S. Govt. Organizations

10 Most Circulated

6 Dr. Melissa, Wikipedia

14 Snopes.com



Post
Collection

- Collected Posts using FB Graph API
- Data Collected From January 1, 2015 – April 2, 2018

Article
Collection

- Used Python Package to collect Article Content
- 29,047 articles from reliable and 15,017 articles from unreliable outlets

Filtering

- Identified Topic by Google Cloud NLP API
- Filtered out articles not related to health



18,436

+

9,153



27,589

Final Corpus

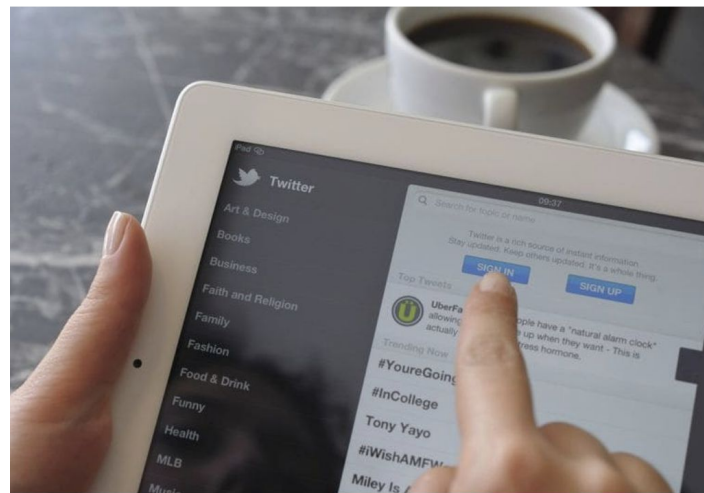
What's in the Headline?

A longer headline receives more click than a short line does. [Breux, C. (2015)]

- Unreliable outlets (**12.13 words/headline**) use longer headlines than reliable outlets. (**8.56 words/headline**)
- An unreliable outlet's headline **has a higher chance of receiving more clicks** or attention than a reliable outlet's headline.

Internet Culture

6 in 10 of you will share this link without reading it, a new, depressing study says



(iStock)



By **Caitlin Dewey**
Reporter

What's in the Headline?

Examined the clickbaitiness of the headlines.

- Used two supervised clickbait detection model (Cohen's $\kappa = 0.44$)
- Considered headline as a clickbait if both models labeled it as clickbait.
- **Unreliable outlets (40.03%) practice more clickbait than reliable outlets (27.29%).**

Internet Culture

6 in 10 of you will share this link without reading it, a new, depressing study says

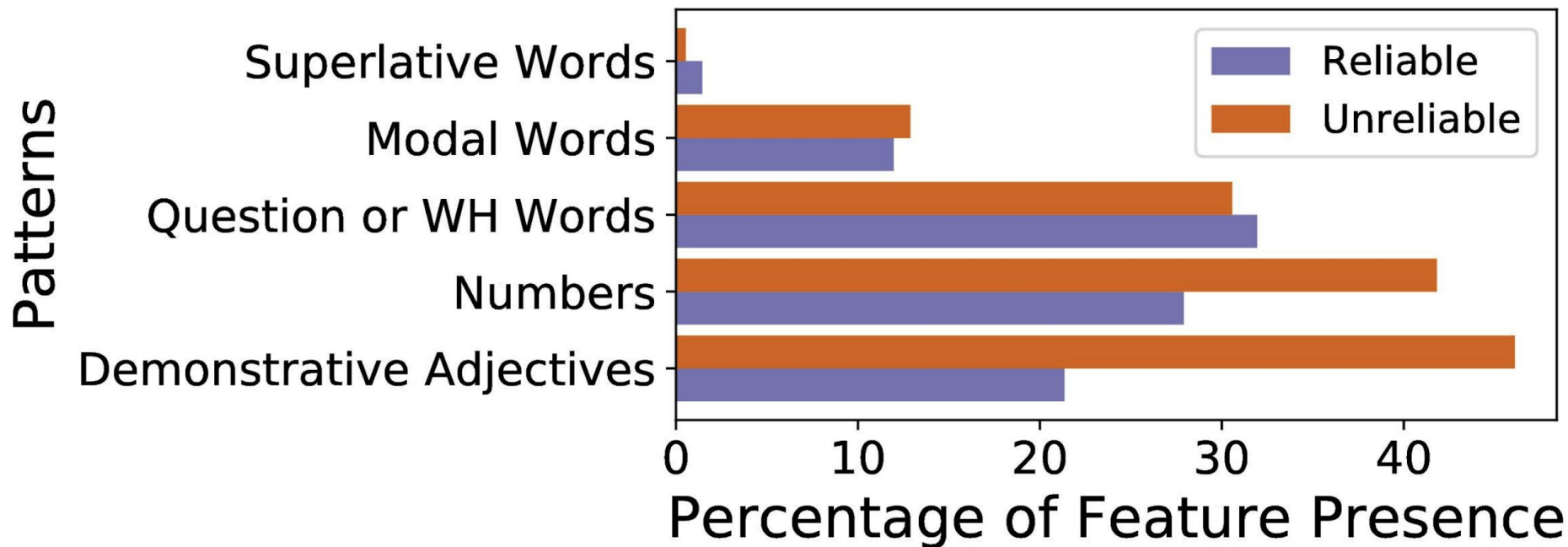


(iStock)



By **Caitlin Dewey**
Reporter

What's in the Headline?



Is everything you think you know about depression wrong?

In this extract from his new book, Johann Hari, who took antidepressants for 13 years, calls for a new approach



Insomnia in adults linked to heart attack and stroke

By Sandee LaMotte, CNN

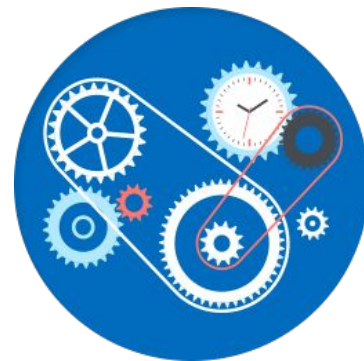
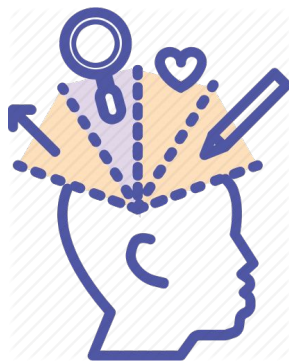
Updated 5:17 PM ET, Wed November 6, 2019



Research Question: Which headlines work best to engage audiences to reliable health-related information.

High Level Goals

1. **Find out the engaging headline forms**
2. Examine readers' perception about these headlines
3. Generate automated headlines



1. Engagement Analysis

2. Perception Analysis

3. Generate Headline

Headline Forms

Going over the headlines, we identified two characteristics

- Informativeness: *How many WH questions are answered in the Headline*

1 WH: You won't believe what Ebola is!

2 WH: Vegetarians might have higher risk of stroke than meat eaters, study says

3 WH: A third person died of EEE, a rare mosquito-borne illness, in the US

Headline Forms

Style = *Advisory, Question, Modal, Superlative, Listicles*

Advisory: *You're literally eating microplastics. How you can cut down exposure to them.*

Question: *Do pregnancy and childbirth accelerate aging in women?*

Modal: *What you should know about cataract surgery*

Superlative: *Best and worst thing to eat to stay healthy*

Listicles: *10 things you need to follow to be healthy*

Data For Preliminary Study

Three news source. 20 health articles from each.

- CNN and Fox: most watched cable news
- Wpost: 5th most circulated newspaper
- Why not New York Times?
 - Requires subscription
 - Convenient sampling
- Why not USA Today [the most circulated]?
 - Health page doesn't give access to past news articles

Ongoing Data Annotation

CNN

Highly Informative [3+ WH questions]	15
Medium Informative [2 WH questions]	4
Low Informative [1 WH question]	1
Question	1
Advisory	8
Listicles	1
Modal	1

Research plan

An online survey with 20 university students would be conducted for the pilot study

Incorporating the lessons from the pilot, the main survey would be conducted with 500 people using Amazon Mechanical Turk to test which headline forms are most engaging.

Considering survey findings, automated headlines would be generated.

Automatic Headline Generation

- Challenges:

- Transferring most engaging styles to the generative model
- Ensuring generated headline will not be misleading
- Lack of available dataset which contains different style information

- Current Approach

- Running experiment with the news articles collected from Clickbait Challenge*
- Developed a GRU (Gated Recurrent Unit) based sequence encoder-decoder along with an attention layer [Bahdanau et al. 2014] to build the model
- Trained the model by randomly selected 5K articles from the dataset and feed first 200 words as an input and the headline as output

- Future Approach

- Exploring Pointer-generation network for headline generation [See et al. 2017]
- Recurrent Neural Networks (RNNs) based Variational Auto-Encoder(VAE) [Shen et al. 2017]

*<https://www.clickbait-challenge.org/>

Potential Impact

- The study would contribute to increasing engagement with reliable health-related information
- Results could be helpful for health agencies like the Centers for Disease Control and Prevention (CDC)
- Its results also would potentially be applied to other domains like environment, politics, and so on.

References

1. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
2. See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. arXiv preprint arXiv:1704.04368.
3. Shen, T., Lei, T., Barzilay, R., & Jaakkola, T. (2017). Style transfer from non-parallel text by cross-alignment. In Advances in neural information processing systems (pp. 6830-6841).

THANK YOU

ANY QUESTIONS?

Contact

Naeemul Hassan (nhassan@umd.edu)