7/23/2017

**User Evaluation Report**

**Team Correct**

Team Members:

## Michael Brennan (16200629)

## Kunal Chavan (16204154)

## Ketaki Ghatage (16200153)

## Xiaosheng Liang (15211913)

Jerom Joy Vattakeril (15200123)

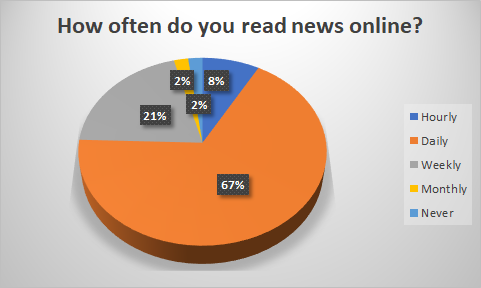
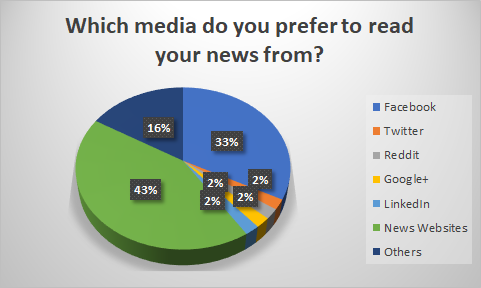
# 1. Proposed Hypothesis (What Questions Are You Asking?)

Explain the questions you are asking in your experiment and why they are important. Also provide an overview of some interesting, important, and relevant existing work in relation to evaluating such hypotheses (e.g., how is this hypothesis typically evaluated, do previous evaluations, if any, have any flaws?). For existing work provide appropriate citations and screenshots.

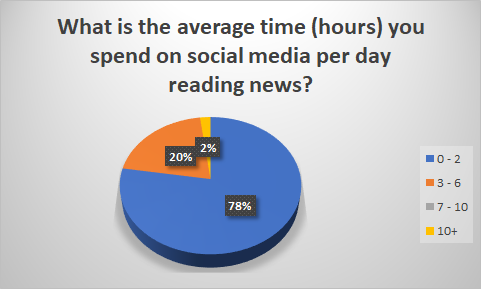
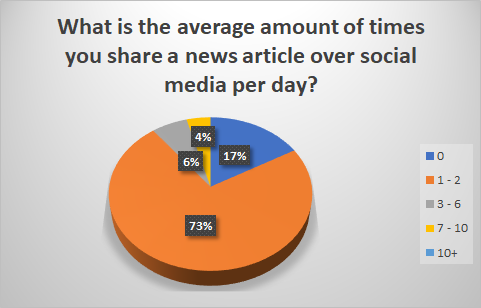
Hypothesis: “People while checking a news article on fake news detection website are most likely to prefer a response in real time and a detailed explanation rather than one word response”

In today’s digital world fuelled by social media, spread of fake news and alternative facts is on rise. So, the need for people to educate about this is more than ever before. Further they need to be empowered with a tool to safeguard themselves. Fake news detector websites are one of the best tools to prevent spread of rumours, but the display of results in real time should be one of the major features to help prevent this spread. As our competitors do not have this feature and cannot provide more detailed explanation for their declaration of the news type, we analysed the features missing in such tools and are implementing it in our system.

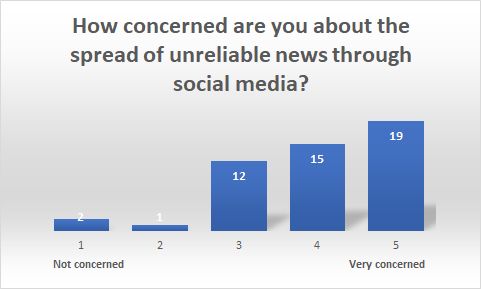
Before going ahead on this, we thought it was necessary to get an understanding of people’s habit of accessing news, their opinion about existence of fake news, and further understanding their demand for a tool to prevent this. For getting this information, we conducted an online survey where we tried to understand the frequency of reading news, media sources preferred by people, their opinion on spread of unreliable news through social media and their concern about this issue. We got responses from 49 users and results are below.

From above results it is clear that people read news online daily and more importantly they are preferring mainly social media sites followed by official news websites.

From above results it is clear that at least 22% of the users spend more than 3 hours per day reading news through social media. Further, at least 10% of the users share above 3 articles every day, and that’s quite high. If they are following sources with fake news, that can have a huge impact on other people following them.

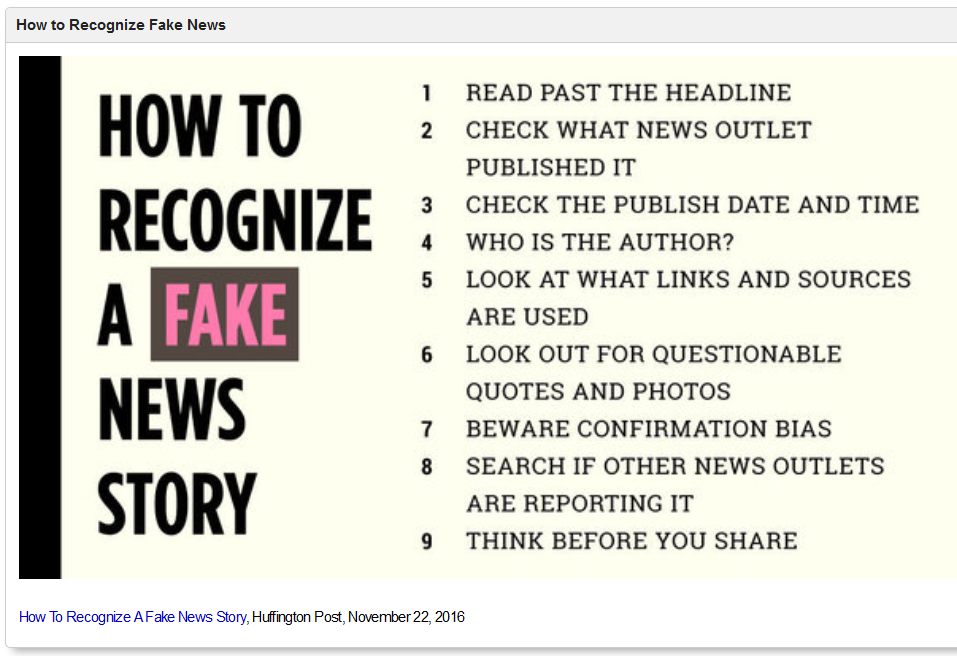
 

Above results show that most users are now aware of the existence of fake news and are very concerned about the same. Many even expressed need for a tool to prevent this.

While checking for such existing tools we came across some famous ones like snopes.com and factcheck.org which analyse each news item manually and hence even though accurate, cannot provide real time results.

Some other tools like bsdetector.tech and fakenewsai.com have avoided any manual investigation, hence are faster. But they primarily depend on a manually compiled list of domains with references to unreliable sources, and this is not enough as everyday new sources come up and it is important to investigate more than just on basis of a domain list.

On further research, we found guidelines towards recognizing fake news items from various reputed sources like [Huffington Post](http://guides.lib.berkeley.edu/c.php?g=620677&p=4328357), [Facebook](http://uk.businessinsider.com/facebook-how-to-spot-fake-news-2017-4?r=US&IR=T), [Harvard Library](http://guides.library.harvard.edu/fake) [1][2][3]and [Factcheck.org](http://www.factcheck.org/2016/11/how-to-spot-fake-news/)[4]. An example given below.



*Figure 1How To Recognize A Fake News Story, Huffington Post, November 22, 2016*

Based on this study made, we understood the need for a fake news detector website which provides the results in a detailed format with information proving the point.

Adding on to the approach of other tools, and using some of the guidelines from above sources (like date of release, presence of author, type of sub-urls used, correlation between title and content, etc.), we came up with a website where instead of using domain, we have taken up a generalized approach and are analyzing entire article’s content and scoring each parameter for more accurate results. Further these scores would be aggregated and displayed to the users.

As this is a fresh approach towards this issue, user evaluation experiment had to be newly designed as well. Our main focus is on educating users towards issue related to fake news and understanding their point of view and expectations from the system. We have designed the experiment to evaluate interface as well as content but mainly focusing on the latter. It is important for us to provide a proper justification to the users and not just a simple indicator about the news articles they want to check, hence main focus on that.

For evaluating the interface, we have used [general guidelines](http://www.ed.ac.uk/information-services/library-museum-gallery/finding-resources/library-databases/databases-overview/evaluating-websites)[5] available and for evaluating the content we have added further to the general guidelines related specifically to fake news detection websites.

# 2. Experimental Method

## 2.1 Overview

How will you run your experiment? What will the experimental conditions be? What is your overall experimental design? What are the baselines you compare against and the evaluation metrics?

Our experiment will be classroom based where participants (would be attracted using advertisements (ex. posters) and some complimentary goodies. Participants) will evaluate our website in comparison with one more similar website, i.e. fakenewsai.com. We have chosen this website as it is uses a very similar approach, i.e. it accepts article url from users and returns a result in real time, unlike other well known fake news detection websites like snopes or factcheck which work on each article manually and take more time to process each article. Evaluation will involve checking news article urls on each website. Initially, demo urls would be given for each real and unreliable articles so that participants would get a better understanding. Further they will be asked to check three more urls of their choice. In case participants do not have any ready, backup urls would be provided.

While evaluating participants would be introduced to websites as System A and System B instead of using actual website names. We will mask the websites so that participants won’t be able to see website logo or any other information that could lead to disclosure of website details. The reason for doing this is to avoid any kind of bias possible [6] as it may affect the result of the experiment.

Once they have completed evaluation for each website, they will be asked to complete an online survey which would mainly focus on understanding participants’ experience while using both the websites (System A and System B). Getting their opinion related to ease of use, speed, appearance (for both homepage as well as details page). Also, comparison between information provided about the articles like if it’s a one/two word response or a detailed explanation and an appropriate score assigned to it.

## 2.2 Data Collection

What data will you collect during the experiment about participants’ performance and why? Will there be a mix between subjective (qualitative) and objective (quantitative) measurements? How does the data you will collect relate directly to your experimental hypothesis?

During our experiment, we will collect data based on participants’ experience while using our website in comparison with another similar fake news detection website (fakenewsai.com). We will mainly use participants’ views on website’s ease of use, speed, and interface. Also, getting to know about participants’ satisfaction with explanation provided about the article and scoring assigned for each parameter, and if they thought it was reasonable enough in comparison with one or two word indicator from other website.

Further the participants would be asked to complete an online survey, where they would answer questions about various parameters for both the websites. Questions would be like:

- How likely are you to return to this website?

- How likely are you to recommend this website?

- Did you experience any problems while checking the URL?

- Did the layout of details page please you?

- Are you pleased with the explanation provided on details page?

This will help us understand how participants felt about our website in comparison with another similar website.

The measurements would be mainly subjective as we are depending on participants’ experience using the websites. We will collect data related to design of website, how appealing was the homepage, how simple and clear was the details page, how effective they found the explanation provided in comparison with one or two word answer, and how effective and useful was the scoring done for each parameter. Also, the data collected from the survey can be used for further improvement in terms of layout design, accuracy of results, etc.

## 2.3 Selected Subjects

Who will you use as subjects in your experiment. Why are these a representative sample? How will you source these subjects?

As news is read by people of all age groups, we preferred to consider age 18 and above. The team discussed and decided that we would not consider people below age 18 as we think they would not understand the importance of the survey and would provide random feedback which would end in incorrect results. This website is developed to verify the English language based news and hence our audience would be people who read most of the news through social media.

The user feedback will be collected from people of different nationalities having knowledge of various news websites. They would also be part of the practical setup by voluntarily contacting us through email and must be quite active on social media. Posters would be put up around college and the practical setup would be conducted during a set period of time. In case we face less turn out for the evaluation we plan to have friends and family taking part in this survey. The practical setup would be conducted offline in the classroom and the survey would be shared online.

## 2.4 Data Analysis

How will you analyse the data you collect during the experiment? How will this analysis answer the question you originally proposed?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Response no.** | **Question 1** | | **Question 2** | | **Question 3** | |
| **System A** | **System B** | **System A** | **System B** | **System A** | **System B** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Once the questionnaire has been completed by the participant, their response would be noted in above format. Here for each question, participant will answer for System A in comparison with System B or vice-versa.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response no.** | **Question 1 Preference** | **Question 2 Preference** | **Question 3 Preference** | **Overall Preference** |
| **System A or System B?** | **System A or System B?** | **System A or System B?** | **System A or System B?** |
|  |  |  |  |  |
|  |  |  |  |  |

Once, the individual responses are collected, they can be aggregated on basis of the question. For instance, if we consider the question ‘How likely you will return to the website?’ For this question, we have options ‘Very likely, Likely, Unlikely and Very unlikely’. So whichever system gets more positive response, that system will be the preferred one for that particular response.

This data can be further simplified once we combine it question wise and get aggregate value for each system. This can be very useful in creating visualizations for each question as well as for overall questionnaire.

This analysis will help us understand participants’ feedback about each parameter for both the websites as each question represents one. Like ease of use, interface, speed, contents on details page, ex: detailed scored response provided or simple one word response, all can be analysed separately. Once combined, it can be used to get an overall result which is our main aim, to compare our website to a similar other fake news detection website.

## 2.5 Practical Setup

In practical terms how will you run your experiment? Will it be online or offline? What instructions will participants be given? What type of room/environment will the experiment be in? Will questions/surveys be displayed on-screen or on

printed paper?

The participants evaluation would take place by meeting the participants in the classroom on set specific dates to maintain the same environment throughout the user evaluation. Information about two websites would be provided and the motive behind building this system shall be explained. We intend to display 2 different websites to our user and hiding the website name, the user would be asked to use both the systems with the URL’s they choose from social media. Two URL’s would be provided to the user to help them understand the working of the system where one is a demo and the other is for the hands-on experience. Each participant is given approximately 10-15 minutes for the entire evaluation.

The survey would be conducted offline on July 28th and August 1st in B.106 Activity classroom. As this classroom is booked for project and we would have easy access. As 28th is Friday, we may have low turnout of participants hence we decided to conduct the survey on August 1st as well. The feedback questionnaire consisting of around 15 – 20 questions is provided online on our system. Each user would be provided one/ two computers to run this test. The number of systems depends on the crowd flow. The users would be instructed on the entire system and what is the goal.

This survey would ask questions about the speed of the system, User Interface, etc. The survey would also ask questions about how the scoring feature of the system as it is a major feature which stands out from the other available websites. Scoring is done based on the content, title, author, date, etc. The feedback would be collected during the team’s absence hence giving the user the freedom to rate as per their real experience.

# 3. Conclusions

What can you conclude after running this evaluation study (e.g., will you learn anything useful about the quality of your solution)? What could be done better in a future evaluation study (e.g., would you like to ask different questions or use different evaluation metrics)?

The user evaluation process would help us in understanding our system in detail from the user’s point of view. We cannot rely completely on user testing, as the chances of missing out on some features are high. As we are focussing on quality of the feedback than quantity, we hope to receive actual feedback from users and this would help us to improve the system. The study results will help us in understanding the real requirements in the market and how our system requires changes based on the demand on comparison to our competitors.

As scoring feature is completely new for this type of website or software, we hope to get more comments on this. The comparison of the two systems, would help us in understanding the positives of the other system and give us idea to build the next version for our website.

The evaluation process is an interactive one as we meet the user’s face to face on campus and explain the concept. We could improvise by working on suggestion box, where the pen and paper format is used to obtain the information and in detail experience from user rather than few limited options. As we believe each user will have a different perspective to the website, understanding their explanation and knowing the implementation feasibility for the website’s next version would be a major learning. We also will prioritize the major features which we feel would require importance and have more questions on those features.

# References

**[1]** *Real News/Fake News: Detecting Fake News.* (22 November 2016). Retrieved from <http://guides.lib.berkeley.edu/c.php?g=620677&p=4328357>

**[2]** *Facebook will now teach you how to spot fake news* by Rob Price, Business Insider UK. (12 April 2017). Retrieved from <http://uk.businessinsider.com/facebook-how-to-spot-fake-news-2017-4?r=US&IR=T>

**[3]** *Fake News, Misinformation, and Propaganda* by Harvard Library. Retrieved from <http://guides.library.harvard.edu/fake>

**[4]** *How to Spot Fake News* by [Eugene Kiely](http://www.factcheck.org/author/eugene-kiely/) and [Lori Robertson](http://www.factcheck.org/author/lori-robertson/). (18 November 2016). Retrieved from <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>

**[5]** *How to evaluate website content* by The University of Edinburgh. (4 July 2017). Retrieved from <http://www.ed.ac.uk/information-services/library-museum-gallery/finding-resources/library-databases/databases-overview/evaluating-websites>

**[6]** Anthony J. Viera, Shrikant I. Bangdiwala. (2007). *Eliminating Bias in Randomized Controlled Trials: Importance of Allocation Concealment and Masking.* Research Series, Vol. 39, No. 2, p132-137.

**[7]** Selecting Subjects by [Dr. Christopher L. Heffner](https://allpsych.com/author/apsupport/). Retrieved from <https://allpsych.com/researchmethods/selectingsubjects/>

**[8]** *Information about qualtrics survey tool.* Retrieved from <https://www.qualtrics.com/>

**[9]** *Planning and conducting survey.* Retrieved from <http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/5-planning-and-conducting-survey>