



## EFFECTS OF HUMAN AND HUMAN-MACHINE FAKE NEWS DETECTION APPROACHES ON USER DETECTION PERFORMANCE

Efeosasere Moibi Okoro  
Department of Computer Science  
National Institute of Construction Technology  
Uromi, Nigeria

Benjamin Abara  
Department of Computer Science  
National Institute of Construction Technology  
Uromi, Nigeria

Aneyelewa Alan-Ajonye  
Department of Social Science  
National Institute of Construction Technology  
Uromi, Nigeria

Zayyad Isa  
Department of Computer Science  
National Institute of Construction Technology  
Uromi, Nigeria

Alex Umagba  
Department of Computer Science  
National Institute of Construction Technology  
Uromi, Nigeria

**Abstract:** The detection of fake news is a relevant problem-solving mechanism. Literature presents us with three methods of detecting fake news; Human-based, Machine-based and the Human-Machine (hybrid) method. There are questions about which detection approach may be best in detecting fake news. In this study, we test the effects of using two existing methods to detect fake news: Human-based and the Human-Machine approach. Participants performed a small classification task where they were asked to determine if a news article was fake or not. The study used two levels of a within-subject design; where each participant used both the human-based approach and the human-machine approach. Performance and User experience were the dependent variables. Results of the study show that the Human-Machine approach improved user fake news detection effectiveness by 26%. The results suggest that augmenting human intelligence with machines has benefits in fake news detection.

**Keywords:** Fake news detection; Human-Machine Collaboration; Detection Performance; Fake news; Human Factor.

### I. INTRODUCTION

This present study aims to test the impact that different fake news detection approaches have on performance and user experience in determining the likelihood that a news article is fake.

The detection of fake news is a relevant problem-solving mechanism. Social media platforms over the years have been the main means used in spreading misinformation [1]. Research has shown that over 60 percent of American adults rely on the news they read on social media [2] which posed a problem that prompted this work and even when users recognize that there are the possibilities of fake news on social media, they still consider social media applications like Facebook as their source for reliable news compared to traditional media like newspaper.

### II. REVIEW OF LITERATURE

Literature presents us with a range of fake news detection approaches to help users distinguish between fake and real news. These detection approaches can be divided into a human approach, a machine approach and a human-machine (hybrid) approach. But, there are questions about which approach are more effective and provides the best

result in determining the likelihood that a news article is fake.

#### A. The Human Detection Approach

The human detection approach is humans deciding without the aid of a tool if a news article is fake or not. This is usually a fast method of classification but, the limitations of this approach are; humans cannot adequately discern between real and fake news. Research shows that humans are only 54% effective at doing so [3] and these could be because of a wide range of reasons such as the time to compute the information, the lack of information available on the subject matter, the lack of understanding of the material and illusory-truth effect.

##### 1) Time to Compute the Information

Humans cannot fact-check the enormous volume of information that is generated online [4 - 6].

##### 2) Lack of Relevant Information Available

Fake news are persuasive and the processing of persuasive information involves absorbing, interpreting and evaluating information [7] and people do not take efforts to probe information.

### 3) *Comprehension*

Humans often lack the required literacy skills to critically examine and discern if a news article is fake or not [8].

### 4) *Virality*

Humans have the illusory-truth effect meaning that the more the visibility of a news article, the more its believability [9 - 10].

Studies have shown that there are two methods in which humans process persuasive information: systemic and heuristic processing [7].

The systematic processing or central route method involves careful consideration and scrutiny of the content of the information and the idea or motive behind it. When humans have ample knowledge on the subject matter coupled with much time and motivation to ascertain the validity of a news article, they tend to perform better in producing more accurate discernment of the information [11-12]. Although this route proves somewhat effective in detecting fake news, however it is rather time consuming and requires more information processing capability.

The heuristic processing or peripheral route method, relies on the use of mental shortcuts or cues to discern the validity of an information. This uses simple rules of thumb such as the belief that professionals or respected people do not lie and their information is authentic and can be trusted or simply because the idea behind the information makes one feel good, the information has to be true. This method although a very fast approach, but according to the Elaboration Likelihood Model theory, this approach sometimes relates to the fact that the user may have little information on the subject matter or lacks motivation to conduct a more cognitive processing and this often leads to a biased judgement.

## B. *The Machine-based approach*

To solve the drawbacks of the human-based approach, the machine-based solutions were employed. The machine-based detection approach has broadly been categorized into two; Corpora / Linguistic method and the Network analysis approach [4].

### 1) *Linguistic approach*

This can also be called the corpora method. This approach considers the textual content of the news article. The content is extracted and analyzed to associate language patterns with deception. This method has proven effective according to [13]. But the limitation of this approach is predicting sentiments, the prediction of sarcasm and the analysis of unknown words.

### 2) *Network Analysis Approach*

This method relies on the examination of existing body of collective human knowledge to assess the likelihood of new statements to be false [4]. The method goes beyond the analysis of the questionable content itself to collect and compare a wide range of similar and related statement from various sources (network) such as metatags like URL and social network behaviour to ascertain the likelihood of the content being false.

The drawback of the URL based detection approach is that there are no definitive list of fake news websites, no

standardized fake news database or a large-scale benchmark dataset [14-15] and there might be disagreements on the inclusions or exclusions of some websites. The corpora-based method also has its limitations as they require a pre-existing knowledge-base to perform.

## C. *The Human-Machine Approach to Fake News Detection*

With the limitation of the existing approaches discussed in sections 2 and 3 (human-based and machine-based approach), studies and observations by [4] [13] [16-18] have shown that there are benefits to the usage of a hybrid approach that combines the strengths of the human approach and machine approach for fake news detection.

[16] theorized a novel approach for fake news detection on Facebook that combines the efforts of machines and the human based social media news literacy education tool. The approach generates a fake news likelihood score from 0% to 100% based on the 10 factors outlined by the social media news literacy education tool by [19]: heading, URL, source, formatting, photo, date, evidence etc. with inputs involving both machine approach such as database checking, URL analysis etc. 0% meaning no likelihood of a news article to be fake and 100% meaning certainly that the news article is fake. Other interesting human-machine approach are those provided by [20] and [18] for detecting fake news on twitter using crowd sourcing.

## D. *Hypothesis*

From sections 2 and 3, we predict that given the benefit of the Human-Machine approach to fake news detection, humans will exhibit better performance and experience using the hybrid approach as opposed to working solely on their own (table 1).

In order to test the hypothesis, a study was set up with fake news detection approach as the independent variable (with two levels; human approach and hybrid approach) and performance and experience as the dependent variables. Because this is basically a classification task, the performance measures were effectiveness (classification accuracy) and efficiency (time taken), while the user experience measures were cognitive load and confidence. Participants performed a small information classification task with a "ground truth" collection of fake news where they were asked to determine how likely a news article is fake while working solely on their own (H) and with a prototype of the [16] human-machine model (HM).

| Variable type   | Dependent variable | Predicted hypothesis direction<br>*HM<br>**H |
|-----------------|--------------------|--|
| Performance     | Effectiveness      | HM > H                                       |
|                 | Efficiency         | HM > H                                       |
| User experience | Cognitive load     | HM > H                                       |
|                 | Confidence         | HM > H                                       |

Table 1: Shows the hypothesis and its direction.

\* HM stands for human/machine approach \*\* H stands for the human approach.

### III. METHOD

#### A. Study design

The study used an independent measure design with one independent variable: fake news detection approach (with two levels: human-based approach or human-machine approach). Participants were asked to determine if a news article is fake. A within-subject (or related sample) design was used; where each participant used both the human based approach and the hybrid approach. The order of approach presented to the participant was allocated randomly to reduce order effect. The dependent variable was a set of performance and user experience variables. Since it is basically a classification task, the performance measure were effectiveness and efficiency while the user experience measure was cognitive load and confidence.

#### B. Study setting

The study was done in a section of the Information and Communication Technology (ICT) office in National Institute of Construction Technology. Participants were secluded in a conducive environment and were presented with the material for the study (a laptop). The researcher was also in the secluded room sitting adjacent to the participant's right.

#### C. Means of Measurement

##### 1) Effectiveness

Here, performance is measured as the effectiveness (or accuracy) of the classifier. Accuracy is described as the proportion of correctly classified items [21]. It is the percentage proportion of the correctly classified news items and the total document available in the document population (equation 1). At the end of the task, the classifier effectiveness is calculated and recorded.

$$Accuracy = \frac{\sum Cci}{\sum Ddp}$$

$\sum Cci$  = Total number of correctly classified items

$\sum Ddp$  = Total number of documents in the document population

Equation 1: Formulae for measuring classifier accuracy

##### 2) Efficiency

Measured as the time it takes the participant to complete the task. At the end of the task, the time taken to complete each task is recorded.

##### 3) Cognitive Load

The amount of mental workload or effort demanded by a task during a specific time duration [22]. Measured as the ratio of the participant's subjective perception of time (estimated time to complete the task) against the actual time taken to complete the task.

Subjective duration is defined as the subjective evaluation of time [23] and objective duration is the objective evaluation of time. In order to access this, a person is asked to verbally estimate the duration of an activity or task [23]. For this study, participants were asked to make a retrospective subjective estimate i.e. they were not informed they will be making a subjective duration estimate till the task was completed. For the duration judgement involving the retrospective paradigm, a value greater than one represents a temporal overestimate which indicates higher cognitive load and a value lower than one represents a temporal underestimate which indicates lower cognitive load [23].

$$Cognitiveload = (Subjectiveduration) / (Objectiveduration)$$

Equation 2: Formulae to calculation duration judgment

##### 4) Confidence

The confidence participants place on an answer is measured by asking them how confident they are in their answers. At the end of the study, the user is asked how confident they are of their classifications using a 7-point confidence Likert scale ranging from absolutely confident to not confident at all. A 7-point confidence Likert scale was used instead of a 5-point confidence Likert scale because according to [24], the 7-point Likert scale provides a more accurate measure of a participant's true evaluation. Also, each unit increase in the number of scale point produced an increase of .2 units in the observed standard deviation [25].

#### D. Instrument

6 ground truth fake news articles were collected from the Snopes website after a search query 'Nigeria' was entered in the search field. Snopes is a definitive fact-checking website for internet sources, rumors and misinformation [26]. The fake news contained a heading, source URL, content and photograph (Figure 1). A PC running Microsoft OneNote was used as the news reading platform because OneNote was easy to prototype a news reading website and to implement [16] human-machine model. The interface for the human-based approach contained the news headline, author, source, photography and the news content (Figure 1) while that of the human-machine approach contained the model's fake news likelihood prediction and every other element contained in the human-based approach (Figure 1). We selected the [16] approach for the study over [20] and [18] because more Nigerians use Facebook which [16] method is based as oppose to twitter which the [20] and [18] methods are based. A questionnaire was designed with variables such as age range, choice of social media and frequency of usage. A 7-point confidence Likert scale that ranged from "absolute confidence" to "Not confident at all" was used to determine their level of confidence on their classification of the news article. A recorder is used to record their verbal protocol during their interview.

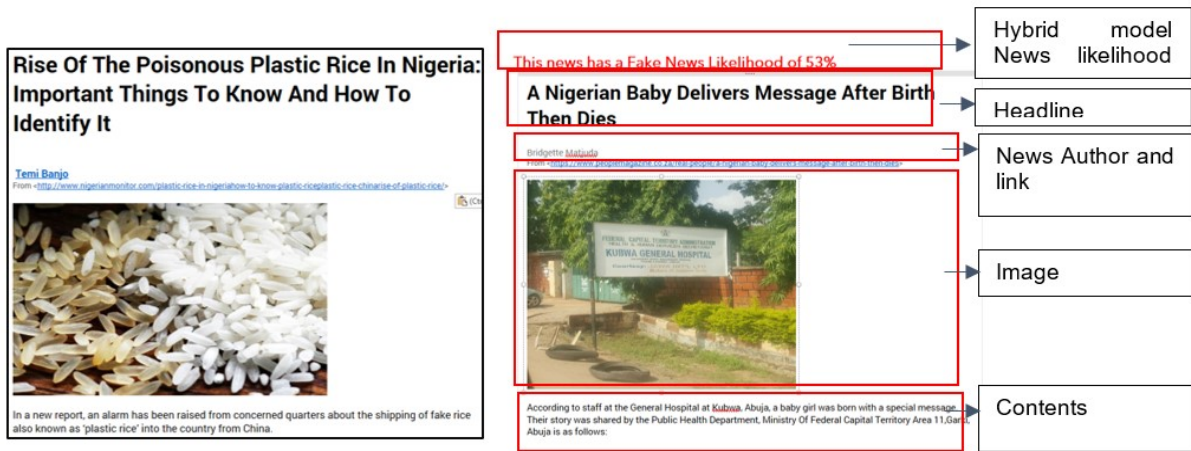


Figure 1: The interface for the human-based interface on the left and hybrid-based approach interface on the right

**E. Participants**

41 participants volunteered for the study. They were all civil servants associated with the National Institute of Construction Technology. All participants voluntarily accepted to take part in the study. 3 were female and 38 were male with an age range of 18 to 45. 10% had their highest qualification as MSc. (Masters), 28% had BSc. (Bachelors), 15% had ONDs (Ordinary National Diploma) and 48% had Senior School leaving certificates (SSCE). 69% read news through social media platforms more than ones a day, 18% at least ones a day, 8% ones a week and 5% ones a month. The language used in the study was English language.

**F. Procedure**

41 participants volunteered for the study but 10 were purposively sampled. They were purposively sampled based on the frequency they read online news through social media, their age and their degree. The condition for being purposively sampled was that a participant (1) read news via social media at least once a week, (2) were at least 25 years old and (3) had at least a Bachelor’s degree.

Participants were given a general training to familiarize them with the news reading platform and the human-machine system. The general training took 2-3 minutes and was done in the same setting as the study.

Participants were then randomly given their first fake news detection approach (either human approach or human-machine approach). For both approaches, the participants were asked to read the news and classify them as fake or real and also determine the percentage likelihood that the news article is fake and how confident they were on their individual news article classification decision. In the hybrid approach, they were told the computer calculate the falsehood likelihood for each individual article.

After the task, they are asked their subjective perception of time (how long they spent on each approach of the study) and they are interviewed to find out reasons for their classifications.

Participants were urged not to discuss the study with fellow participants waiting to take part in the study.

**IV. FINDINGS**

**A. Descriptive Statistics**

Here we show the results of the questionnaire analysis. The questionnaire responses showed that 86% of the participants used social media as their source of news: Facebook 48%, Instagram 21% and Twitter 15% while 16% used other online channels for their source of news such as traditional newspaper websites (Figure 2).

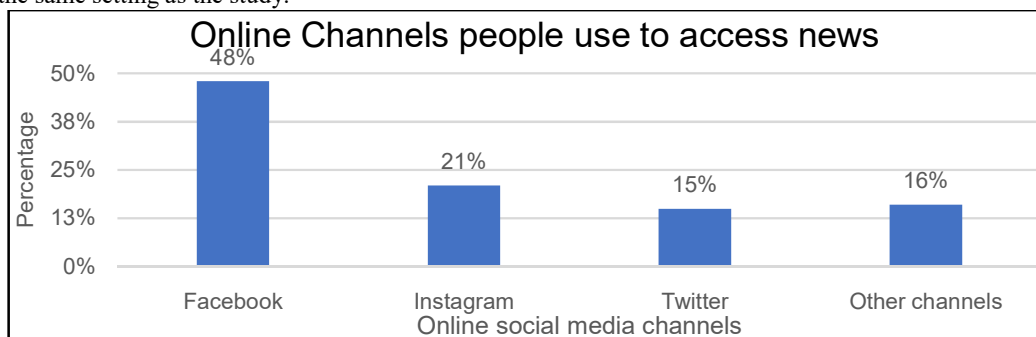


Figure 2 The interface for the human-based interface on the left and hybrid-based approach interface on the right

**B. Inferential Statistics**

Here we show the finding of the study aimed at testing the effects of using different fake news detection approaches on user classification performance and experience. As stated in the method section, we are measuring this effect based on a number of performance and user experience measures: effectiveness, efficiency, cognitive load and confidence.

After the study, A within-subject ANOVA test was performed to test for significance with a 0.05 significance level.

*1) Effectiveness*

Effectiveness is measured as the accuracy of the classifier. Participants using the human-machine detection approach (M=74% SD =0.21) were significantly more effective at classifying news articles as fake than those using the human approach (M =52% SD =0.17),  $F(1,9) = 20.99, p < .05$  (Figure 3).

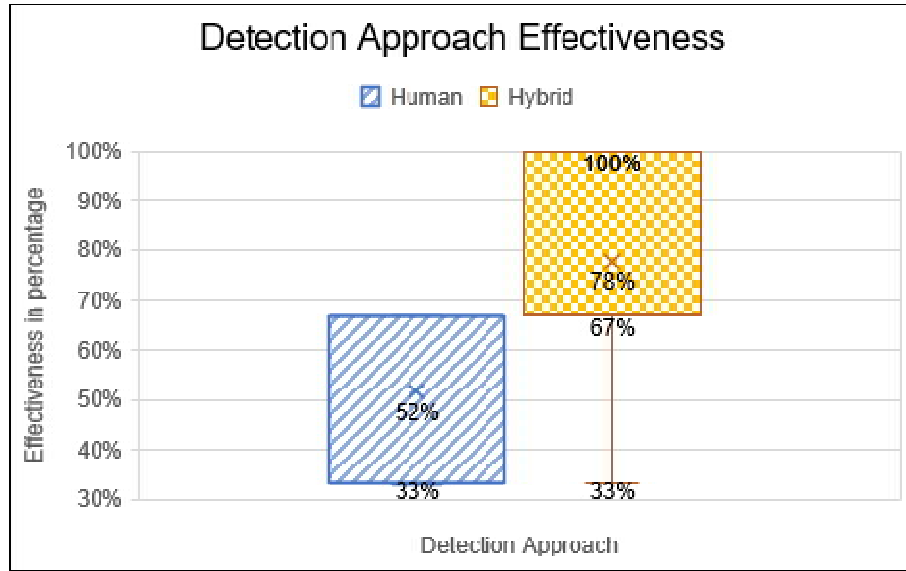


Figure 3: A box and whiskers plot showing the distribution of the effectiveness measure for the human and hybrid detection approaches.

*2) Efficiency*

Efficiency is measured as the time taken to complete the task. The effect of efficiency on participants using the human-machine detection approach (M=6 minutes SD =1 minute 48 seconds) and those using the human approach (M=4 minutes 30 seconds SD = 1 minute 30 seconds) were statistically insignificant,  $F(1,9) = 3.31, p > .05$  (Table 3)

*3) Cognitive load*

Cognitive load is measured as the ratio of the participants subjective perception of time and their objective time taken to complete the task. A value greater than 1 represents a higher cognitive load and a value less than 1 represent lower cognitive load. The effect of cognitive load on participants using the human-machine detection approach (M=1 SD =0.93) and those using the human approach (M=1 SD =0.71) were statistically insignificant,  $F(1,9) = 0.18, p > .05$  (Table 3)

*4) Confidence*

Confidence is measured using a 7-point scale to indicate how confident the participants are of the classification they made from “Absolutely confident” to “Not confident at all”. The effect of confidence on participants using the human-machine detection approach (M =90%) and those using the human approach (M =90%) were statistically insignificant,  $F(1,9) = 0.18, p > .05$  (Table 3)

| Measures   | Hypothesis Direction   | Statistical significant                     |
|--|--|---|
| Effectiveness (percentage)                       | Hybrid (M=78% SD =0.21) > Human (M =52% SD =0.17)  | Significant<br>$F(1,9) = 20.99, p < .05$    |
| Efficiency (minutes)                             | Hybrid (M=6 minutes SD = 1 minute 48 seconds) > Human (M =4 minutes 30 seconds SD = 1 minute 30 seconds) | Not significant<br>$F(1,9) = 3.31, p > .05$ |
| Cognitive load (Perception of time /actual time) | Hybrid (M=1 SD =0.93) < Human (M =1 SD =0.71)  | Not significant<br>$F(1,9) = 0.18, p > .05$ |
| Confidence                                       | Hybrid (M =90%) > Human (M =90%)   | Not significant<br>$F(1,7) = 0.50, p > .05$ |

Table 3: Showing the hypothesis direction and statistical significance of the measure of experiment

## V. DISCUSSION AND CONCLUSION

The aim of this work was to test the effects of human-based (H) and human-machine (HM) based fake news detection approaches on user performance and experience.

Based on the limitations of the human-based and the machine-based approaches, we predicted that users will perform better in terms of performance and experience using a human-machine approach as opposed to a human-based one.

However, the outcome of the study indicated that, the effects of fake news detection approaches on performance (effectiveness) was 26% significantly higher when participants used the human-machine approach as oppose to working on their own. While on the other hand, the effects of fake news detection approaches on user experience (cognitive load and confidence) were not different compared to when they worked on their own.

### A. Increase in fake news detection effectiveness

The 52% detection score by participants using the human-approach confirms [8] study findings that shows that humans are only a little better than chance (50%) at fake news detection. However, the increase in human effectiveness in fake news detection by augmenting human detection skills with machine is a significant finding. According to [17] and [27] humans can perform better when their skills are augmented that of a machine.

### B. Social Media News Readability

We observed that about 70% of the participants visited social media for news daily and 86% of participants rely on the news they read on social media of which the most predominately used channel was Facebook. This finding is inline with [2] research that showed that over 60% of adults rely on social media for their news.

### C. Limitation of the Study

We recognize that there are potential limitations to the study which may have influenced the results. For example (1) the news items were not randomized with real news as the problem set was only made up of collection of fake news. (2) The study participants were asked to pick if news items were fake or not. The method presented them with two choices. During the debrief session. The participants suggested that it would have been ideal if they had a third option (i.e. a maybe option).

In conclusion, this works aim was to test the effects of human-based and human-machine fake news detection approaches on user performance and experience. The study suggested that the fake news detection performance of frequent social media news readers will increase when their efforts and skills are augmented with that of machine-based methods.

## VI. REFERENCES

- [1] C. J. Bond and B. M. DePaulo, "Accuracy of deception judgments.," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 214-234, 2006.
- [2] D. Mrowca, E. Wang and A. Kosson, "Stance Detection for Fake News Identification," Stanford University, 2016.
- [3] E. Mustafaraj and P. T. Metaxas, "The Fake News Spreading Plague: Was it Preventable?," in *Proceedings of ACM Web Science Conference*, Troy, NY, USA, 2017.
- [4] A. Halimeh, P. Pourghomi and F. Safieddine, "The Impact of Facebook's News Fact-Checking on Information Quality (IQ) Shared on Social Media," *MIT International Conference on Information Quality*, 2017.
- [5] N. J. Conroy, V. L. Rubin and Y. Chen, "Automatic Deception Detection: Methods for Finding Fake News," in *ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, St. Louis, MO, USA., 2015.
- [6] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211-236, 2017.
- [7] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer and A. Flammini, "Correction: Computational Fact Checking from Knowledge Networks," *PLoS ONE*, vol. 10, no. 6, 2015.
- [8] R. A. Baron and D. Bryne, *Social Psychology*, London: Allyn and Bacon, 2000.
- [9] M. Gabielkov, A. Ramachandran, A. Chaintreau and A. Legout, "Social Clicks: What and Who Gets Read on Twitter?," *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pp. 179-192, 2016.
- [10] M. Anderson, "Social media causes some users to rethink their views on an issue," 7 November 2016. [Online]. Available: <http://www.pewresearch.org/fact-tank/2016/11/07/social-media-causes-some-users-to-rethink-their-views-on-an-issue/>.
- [11] D. Goodwin-Ortiz, "Fake News on Social Media: Illusory Truth and the 2016 Presidential Election," Thomas Edison State University, New Jersey, 2017.
- [12] Maheswaran and S. Chalken, "Promoting Systematic Processing in Low Motivation Settings: Effect of Incongruent Information on Processing and Judgement," *Journal of Personality and Social Psychology*, pp. 1762 - 1776, 1991.
- [13] R. E. Petty and J. T. Cacioppo, "Involvement and Persuasion; Tradition Versus Integration," *Psychological Bulletin*, pp. 367 - 374, 1990.
- [14] K. Shu, S. Amy, W. Suhang, T. Jiliang and L. Huan, "Fake News Detection on Social Media: A Date Mining Perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.
- [15] W. Hango, "University graduates with lower levels of literacy and numeracy skills," Statistics Canada: Insights on Canadian Society, 2014.
- [16] Okoro, B. Abara, A. Umagba, A. Ajonye and Z. Isa, "A Hybrid Approach to Fake News Detection on Social Media," *Nigerian Journal of Technology (NIJOTECH)*, vol. 37, no. 2, April 2018.
- [17] D. Rus, "More Evidence That Humans and Machines Are Better When They Team Up," *MIT Technology Review: Intelligent Machines*, 8 November 2017.
- [18] S. Karodia, "Annotating the Veracity of Tweets through Mobile Crowdsourcing," University of Cape Town, Cape Town, 2017.
- [19] Facebook, "Tips to Spot False News," 2017. [Online]. Available: [https://web.facebook.com/help/188118808357379?\\_rdc=1&\\_rdr](https://web.facebook.com/help/188118808357379?_rdc=1&_rdr).
- [20] S. Mwanza and H. Suleman, "News Credibility Checking System for Twitter," University of Cape Town, Cape Town, 2017.

- [21] S. Teufel, "AN OVERVIEW OF EVALUATION METHODS IN TREC AD HOC INFORMATION RETRIEVAL AND TREC QUESTION ANSWERING," in *Evaluation of Text and Speech System*, Dordrecht, The Netherlands, Springer, 2007, pp. 168-186.
- [22] R. A. Block, P. A. Hancock and D. Zakay, "How cognitive load affects duration," *A meta-analytic review. Acta psychologica*, vol. 134, no. 3, pp. 330 - 343, 2010.
- [23] M. Sucala, B. Scheckner and D. David, "Psychological Time: Interval Length Judgments and Subjective Passage of Time Judgments," *Current psychology letters: Behaviour, brain & cognition*, vol. 26, no. 2, 2010.
- [24] K. Finstad, "Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales," *Journal of Usability Studies*, vol. 5, no. 3, pp. 104-110, 2010.
- [25] R. Peterson, "A Quantitative Analysis of Rating-scale Response Variability," *Marketing Letters*, vol. 8, pp. 9-21, 1997.
- [26] Snopes, "Snopes," Urban Legends Reference Pages, 2017. [Online]. Available: [www.snopes.com](http://www.snopes.com). [Accessed 31 October 2017].
- [27] D. C. Engelbart, "Program On Human Effectiveness," *Stanford Research Institute*, December 1961.