International Journal of Scientific Engineering and Science
Volume 2, Issue 10, pp. 14-18, 2018. ISSN (Online): 2456-7361

A Model for Visibility Reduction Extended Hybrid Fake News Detection System

Benjamin A. Abara[1], Efeosasere Moibi Okoro[1], Ibe Peace Ibenu[2], Ekene Samuel Nnebi[3]
[1]Computer Science Department, National Institute of Construction Technology, Uromi, Edo State, Nigeria
[2]Mathematics Department, Federal University Lokoja, Lokoja, Kogi State, Nigeria
[3]Computer Science Department, Ambrose Alli University, Ekpoma, Edo State, Nigeria

Abstract*—Since the advent of the internet and mostly the birth of the social media, fake news has escalated significantly and because of this, traditional fact checking approaches such as reviewing each news article has become nearly impossible. Tackling this increase in fake news contents gave rise to the machine-based approach though they faced their limitations which include the absence of a unified database, ambiguities etc. The limitations of the machine-based approach led to the hybrid (Artificial Intelligence-human) approach that combine the efforts of both man and machine resulting to the flagging of news articles as fake. This has shown a lot of promise however; recent research has shown that flagging of news articles as fake has small effect in supporting users in classifying fake news because of a psychological effect called the Illusory truth effect. In this paper, we describe an extended hybrid fake news detection system that not only flags news as fake but also reduces their visibility to reduce Illusory truth effect*.

***Keywords*—** *Fake News, Fake News Detection, Illusory Truth Effect*.

I. INTRODUCTION

The description of what news actually is, is not properly defined. There are so many definitions of the word "News" but the underpinnings of what makes up a news and why it is news is quite vague. It is somewhat agreed upon that News has to be recent and newsworthy, but as to what should be asserted as newsworthy becomes a question. [1] reported that when journalists were asked "how they define news", they sometimes replied with "I know it when I see it". This further brings vagueness as to what qualifies a content to be News. [2] refers to the term News as a primitive construct, one that does not require definition during ordinary conversation, because everyone knows it. According to [3], when journalists are pressed on to say why something has been considered newsworthy, their response is typically "because it just it". [4] further adds that the "just know it" feeling about what News is hides as much as it exposes the values of news selection. This prompted academics to propose their own explanations in form of classifications of news [1]. Hence the question of the "definition of news" was referred to by [1] as a deceptively simple question. This lack of certainty of what news is and its characteristics does not only pose as an issue in news selection by journalists but also has somehow led to a bigger problem called fake news. Since news cannot be defined so fake news cannot be easily defined. But however, based on research, what makes fake news, fake news is the intension [2].

Fake news has grown over the years with the help of the internet due to the ease at which information becomes viral. The internet has created an enabling environment for victims, malicious sharers and engagement-optimized algorithms to share fake news contents that will immediately reach millions of users [3]. The advent of other technologies such as clickbait, applications for generating news, application for photo manipulation, application for video creation tools using AI and 3D modelling [4] and social media platforms aided the virality of fake news. Research shows that trillions of contents are generated by social media users per second and this has led traditional fact checking and news verification processes (such as source verification and the monitoring of every user-generated content shared through social media) inadequate in solving the issue [5-7].

A. Fake News Impact

Fake news has negatively affected several aspects of our lives and its effects includes delayed action, loss of lives, crash of economies, racial, tribal and religious wars etc. in Nigeria and in the whole world [8-10]. A good example is its impact on the Nigerian Federal Government. There was a delayed response of the Nigerian Government towards the Chibok girls' tragedy which was epitomized by the slogan #BringBackOurGirls worldwide. As a general consequence of fake news, the Government claimed the news was a political stunt and hence called the crime a hoax, thereby confusing and delaying efforts to rescue the girls [11]. In 2017, there was the case of a fake news of the death of the Nigerian President, Muhammadu Buhari [12], [16] reported that during the Ebola outbreak in Nigeria in year 2014, because of false publications, citizens resulted to using salt-water to bath as a remedy. Fake news also ripples negative impressions on the religious and ethnic balance in Nigeria. There was a case of fake news that stated Nigeria as the most difficult place for Christians to live, and another fake report that the Nigerian Military are involved in arming and supporting the operations and attack of the herdsmen [13]. These types of news could have profound impacts on politics, society, economy and democracy [2]

B. Fake News Solution

In other to solve this problem many detection approaches have surfaced. Detection Approaches are tools to aid users in the detection of fake news. These approaches are broadly categorized into 3 based on their use and components [14].

They are the Human-based approach, Machine-based approach and Hybrid (Human-Machine-based) approach.

C. The Human-Based Approach

This involves people spotting, verifying and identifying fake news with the help of some guidelines. Traditional news verification process involves journalism processes of investigation and verification [15]. The human-based approach became immediately flawed and tedious with the increase in circulation of fake news contents on social media and thus rendered this approach inadequate [16]. Also, most people lack the skills to critically evaluate and spot fake news [17] because of these limitations, this approach at best have 54% effectiveness [18].

D. The Machine-Based Approach

This involves the use of machines or computational methods to process news contents, transform them into data structures and analyze them to determine the likelihood of the news item to be fake or real [19, 20]. The result was that the machine indicates a news as fake. Artificial intelligence could have been an effective solution but however, this approach faced its limitations such as (1) the absence of a unified database for the Uniform Resource Locator (URL) network-based approach (2) no definitive list of fake news website which also raises the issues of what websites should be listed or added, (3) absence of a pre-existing knowledge-base for a Corpora machine detection approach, (4) inability to model complex dependencies such as semantic relationships [16] and generalization (5) the issue of ambiguity in natural language processing etc.

E. Hybrid Human-Machine Approach

Due to the limitations experienced by both the human-based and machine-based approaches acting individually, the hybrid (human-machine) approach was considered [19]. [18] proposed a human-based and machine based collaborative approach for Facebook using the Facebook education literacy tool. Other researchers such as [25] and [26] have done similar on Twitter that required crowdsourcing.

The [18] model hypothesized a ranking system where both the human-based and machine-based approach will give input towards the ranking system. The model involved the machine likelihood scoring and the human-based method using the guidelines provided by the Facebook's social media education literacy tool called "Tips to Spot False News" provided by Facebook (2017). This tool contained 10 measures to spot fake news on social media. (1) Heading (2) URL (3) New source (4) news formatting, (5) photograph (6) date of publication (7) evidence (8) similar news sources (9) jokes (10) shareability. The ranking was done based on the mathematical equation (MH) equals the summation of all the measures (Equation 1).

$$MH=\Sigma\ (A, B \ldots J)$$
$$Where\ MH \leq 100$$

Equation 1: The mathematical equation for the Human-Machine approach

where

MH is the Machine-Human approach
A - J are the 10 measures.

In an ongoing empirical study, this approach was tested by measuring participants' performance and experience while using both the human-based approach and the hybrid approach. The study showed 26% better performance in detecting fake news than the human approach alone. The users were much more effective in correctly classifying news contents. Approach effectiveness (Classification accuracy) was calculated as the ratio of the total number of correctly classified items (Cci) and the total number of documents on the document population (Dd) (Equation 2).

$$Accuracy = \Sigma Cci / \Sigma Dd$$

Equation 1: The mathematical equation for approach effectiveness

Cci = Total number of correctly classified items
Ddp = Total number of documents in the document population

II. Illusory Truth

Ongoing study results show that there was a 26% increase in effectiveness, but the study did not test shareability and interaction with fake news. More research has shown that flagging and displaying results as fake only has a little effect [21] i.e. they may still share it, may still be biased towards it may still believe it. To fully understand why, literature points to a phenomenon ignored during the design of detection systems and frameworks called the Illusory Truth Effect. This simply means that the more a news content is repeated, the more users are likely to perceive it to be true or valid. The earliest observation of the illusory truth effect was by [28] during a study which noted that subjects rated repeated statements as more probably true than new statements. [29] and [30] also noted that the issue of the illusory truth effects rest on a lot of cognitive attributes such as: recollection, familiarity, fluency, semantic retrieval and misattribution.

Recollection (ability to remember): Recollection is the action of remembering something. A lot of research have been carried out to get the extent of interaction between recollection and the illusory truth effect. [31] reported that people tend to rely on fluency when they fail to recollect credibility of an information's source, hence in the absence of knowledge, illusory truth takes priority. But according to [32] experts experienced increased susceptibility to the illusion, proposing that domain knowledge can hurt rather than help. In addition, [33], described memory as imperfect and it is insensible to trust some remembered facts over another.

Familiarity (false recollection): This is the sense or feeling that an item has previously been encountered without any additional contextual detail about the initial encounter [22]. Users may likely classify a news content as true based on a feeling of previously entering a news content without any evident backing or proper recollection of that event.

Fluency (feeling recognition, consistency): [34] described fluency as the "subjective ease experienced while processing information". Research by [35] demonstrates that fluency can influence people's judgments, even in contexts that allow them to draw upon their stored knowledge. This means that even when users are notified that the likelihood of a news content to be false is high, as long as they see, remotely feel they recognize it, it is easy to process and it is repeated, there

is a high chance it will influence their decision and creep into their knowledge base. This will further lead to referencing that fake news content in conversations and further increase believability.

Semantic retrieval (recollection without source): This refers to the recollection of information where the source is unknown [23]. This occurs when the source of a content recalled is unknown.

Misattribution (attaching recalled information to wrong source): This is the act of attributing a memory or idea to an incorrect source. This happens when a small section of an information is successfully remembered but linked to inappropriate person or time [22]. The means that there is a chance that a false and incomplete recalled information can be attributed to a credible source.

This goes to show that being able to effectively detect fake news is not the end of the problem as users who come in contact with that news content may end up rating it to be true or make reference to it at a later point with a case of misattribution.

The illusory truth effect being the phenomenon where repeated statements has a higher likelihood of being judged true, more research has shown that users can also experience this effect even without repetition. Studies showed that the effect occurred when aphorisms that rhymed received higher truth ratings than those that did not rhyme and statements that contained fonts with higher contrast received higher truth ratings than those of lower contrast [24, 25]. Hence it is believed that fluency maybe the driving mechanism behind the illusory truth effect [26]. Further research also showed that fluent words, names, paintings appear to be more familiar and appreciated. This goes to show that fluency is not just repetitiveness but also the feeling of recognition due to the ease at which it was noticed, like how the eyes are attracted to larger fonts and how the colour red or orange draws more attention.

Hence this paper proposes that a reduction in the fluency of a news article can further reduce the chances of the illusory truth effect and greatly reduce user interaction with that news content. For the cause of this paper, fluency will be measured in two forms of content visibility, (1) size reduction and (2) opacity.

III. Augumenting the Hybrid Fake News Detection Approach with Visibility Reduction Component

This research proposes a system that does not only flag fake news but also reduces the visibility of fake news to decrease Illusory truth effect. This system is an extension of [18] Hybrid Machine-Human (MH) approach to fake news detection by including a visibility component to the system.

A. Description of System

The purpose of this system is to significantly reduce user interaction with fake news contents by reducing its visibility on the social media platform. The rationale behind this system is to implement the reverse of the illusory truth effect. Being that if an information's believability and interactivity increases with when users see it repeatedly, there should be a reverse effect if the information is seen less frequent with little visibility.

The Human-Machine (MH) model will be used to determine and attach a likelihood score to the news article, the likelihood score will be used to determine the extent of visibility attributed to the news content by the visibility component (VC) and finally the visibility attribute will be affected on the social media platform (Figure 1). This means that the HM approach provides the likelihood score to the news content and passes on the score to the VC. The visibility module now assigns a visibility attribute to the content based on the likelihood score provided and then the social media platform now implements this attribute to the news content.

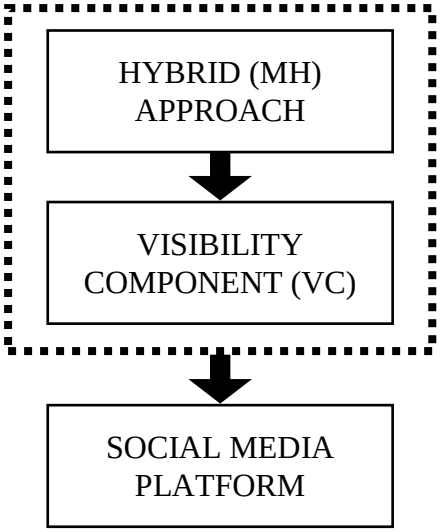


Fig. 1. Showing a schematic description of the Hybrid (HM) approach with visibility component.

$$MHVC = MH + VC$$
$$MHVC \geq 100$$

Equation 3: Relationship between the machine-human model and the visibility component.

Where MH is the hybrid approach and VC is the visibility component. For example, if there is a news article collection N and there are three article n1, n2 and n3. MH reports n1, n2 and n3 as having 100%, 40% and 0% likelihood to be false, n1, n2 and n3 visibility reduces 100% (disappears), 40% and 0% (very easily seen) respectively.

1. Given news article collection N
2. For every news article n
   (a) MH ≤ 100
   (b) F{ no , nz} = MH

Where
MH = Machine-Human Likelihood score
Visibility (Fluency) F = News opacity no and news size nz

Pseudocode 1: Pseudocode showing how the visibility module can work at an interface level

B. Mathematical Model Formulation

In this section, we build a model to depict the interaction with fake news, the subsequent spread of this news due to Illusory truth effect and the incorporation of the proposed solution which is the visibility reduction component.

In this work, important variables governing the Illusory truth effect are: Recollection, Familiarity, Fluency, Semantic retrieval and Misattribution. Each of these variables can be grouped into three categories of influence which is shown below:

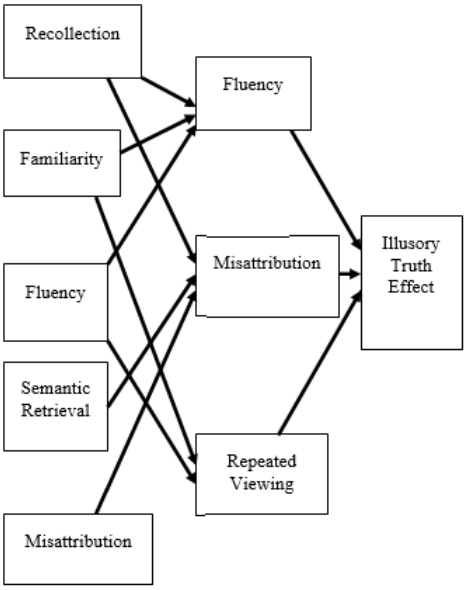
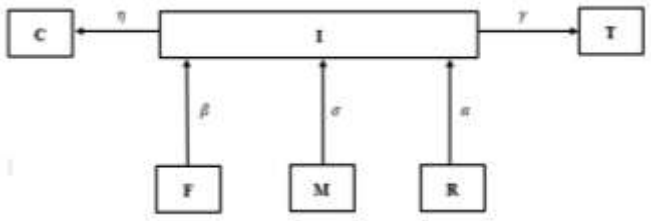


Fig. 2. Showing a grouping of the governing variables.

We develop a mathematical model for Illusory truth effect incorporating the proposed solution of visibility component, font reduction and contrast reduction. We assign the following variables to the various components: Illusory truth effect I(t), Misattribution (M), Fluency (F), Repeated viewing (R), Font Reduction (T), Contrast Reduction (C).

The flowchart for the reduction and control of Illusory truth effect is shown below:



Thus, the mathematical model for the control and prevention of the interaction with, and spread of fake news by individuals is represented by the following differential equation:

$$dI / dt = \alpha R + \beta F + \sigma M - \eta C - \gamma T$$

Here β, σ and α represents the rate at which fluency, misattribution and repeated viewing each, contributes to the Illusory Truth Effect and η and γ each represents the rate of effect which the control measures have on the prevention of the spread and interaction of fake news.

IV. Conclusion

The detection of fake news is a highly relevant problem-solving mechanism [14]. Many detection approaches have been developed but they all ignore the illusory truth effect which means as long as fake news is exposed to people whether they are flagged or not, they will still believe it. There is a need to reduce this illusory truth effect and one way to do so is to reduce the likelihood for them to be seen and shared. Based on this, this paper providing an extension of the hybrid machine-human (MH) approach which include a visibility reduction (VC) approach is necessary. The approach combines the hybrid fake news detection approach by [18] and a visibility component. The system relies on the hybrid approach to provide a likelihood score which then informs the visibility component to reduce the visibility (news opacity and news size) of the fake news.

A. Limitation

The model has a limitation: the hybrid approach by [18] does not provide the number of ratings before it is validated to take on the particular score.

B. Further Work

If we are right that there is a benefit in this relationship between the hybrid approach and Visibility reduction component, then the benefits should be quantifiable or empirical. The measures that could be considered are interaction (in terms of shareability) and believability. Given a dataset of both fake and real news, an empirical study can be conducted to evaluate the extended approach. Based on research we can then provide a hypothesis that users using the visibility reduction component extended approach will share and believe less fake news than users using the hybrid model alone.

References

[1] T. Harcup and D. O'Neill, "What is News?," *Journalism Studies*, no. 12, pp. 1470-1488, 2017.
[2] P. J. Shoemaker, "News and newsworthiness: A commentry," *Communications*, vol. 31, no. 1, pp. 105-111, January 2006.
[3] P. Brighton and D. Foy, News Values, SAGE Publications, 2007.
[4] I. Schultz, "The journalistic gut feeling," *Journalism Practice*, vol. 1, no. 2, p. 190–207, 2007.
[5] J. Meinert, M. Mirbabair, S. Dungs, and A. Aker, "Is it really fake? – Towards an understanding of fake news in social media communication," in Social Computing and Social Media. User Experience and Behavior, pp. 484-497, 2018.
[6] A. X. Zhang, A. Ranganathan, S. E. Metz and S. Appling, "A structured response to misinformation: Defining," *Web Conference Companion*, 23-27 April 2018.
[7] S. Suwajanakorn, "Fake videos of real people - and how to spot them," 25 July 2018. [Online].
[8] 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.
[9] 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.
[10] 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.
[11] J. H. Brunvand, American Folklore: An Encyclopedia, Taylor & Francis, 1998.
[12] K. Rapoza, "Forbes," 2017. [Online]. Available: https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/#5e66ab9d2fac. [Accessed 06 July 2017].
[13] D. Johnson, V. Nayar, and S. Yadav, "The India WhatsApp video driving people to murder," BBC, 2018.

[14] S. Busari, "How fake news does real harm," 24 April 2017. [Online]. Available: https://ted.com/talks/stephanie_busari_how_fake_news_does_real_harm/.
[15] O. J. Nwachukwu, "Ex-British lawmaker, Eric Stuart pronounces President Buhari dead," 22 May 2017. [Online]. Available: http://dailypost.ng/2017/05/22/ex-british_lawmaker_eric_stuart_pronounces_president_buhari_dead.
[16] Premium Times, "Ebola sparks panic across Nigeria as citizens scramble for salt-water bath "remedy"," 8 August 2014. [Online].
[17] The Nation Newspaper, "Military not arming Fulani herdsmen to kill Christians, says Govt," 7 February 2017. [Online]. Available: Yusuf. A., (2017, February 7), Military not arhttp://thenationonlineng.net/military-not-arming-fulani-herdsmen-kill-christians-says-govt/.
[18] E. 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.
[19] W. Dean, "Journalism as a discipline of verification," 2018. [Online]. Available: https://www.americanpressinstitute.org/journalism-essentials/verification-accuracy/journalismdiscipline-.
[20] M. Mohtarami, R. Baly, J. Glass, P. Nakov, N. L. Màrquez, and A. Moschitti, "Automatic Stance Detection Using End-to-End Memory Networks," North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 2018.
[21] N. Delellis and V. Rubin, "Educators' Perceptions of Information Literacy and Skills Required to Spot 'Fake News'," Association for Information Science and Technology, November 2018.
[22] C. J. Bond and B. M. DePaulo, "Accuracy of deception judgments," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 214-234, 2006.
[23] S. Tschiatschek, A. Singla, M. G. Rodriguez, A. Merchant, and A. Krause, "Fake News Detection in Social Networks via Crowd Signals," *Companion*, 23-27 April 2018.
[24] N. Roberto, "Word sense disambiguation: A survey," *ACM Comput. Surv.*, vol. 41, no. 2, 2009.
[25] S. Mwanza and H. Suleman, "News Credibility Checking System for Twitter," University of Cape Town, Cape Town, 2017.
[26] S. Karodia, "Annotating the veracity of tweets through mobile crowdsourcing," University of Cape Town, Cape Town, 2017.
[27] G. Pennycook and D. Rand, "The implied truth effect: Attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings," Social Science Research Network, 8 December 2017.
[28] L. Hasher, D. Goldstein, and T. Toppino, "Frequency and the conference of referential validity," *Journal of Verbal Learning*, pp. 107-112, 1977.
[29] J. P. Mitchell, C. S. Dodson, and D. L. Schacter, "FMRI evidence for the role of recollection suppressing misattribution errors: The illusory truth effect," *Journal of Cognitive Neuroscience*, vol. 17, no. 5, pp. 800-810, 2005.
[30] J. P. Mitchell, L. Sullivan, D. L. Schacter, and A. E. Budson, "Misattribution errors in Alzheimer's disease: The illusory truth effect," *Neuropsychology*, vol. 20, no. 2, pp. 185-192, 2006.
[31] C. Unkelbach and C. Stahl, "A multinomial modeling approach to dissociate dfferent components of the truth effect," *Consciousness and Cognition*, vol. 18, pp. 22 -38, 2009.
[32] H. R. Arkes, C. Hackett, and L. Boehm, "The generality of the relation between familiarity and judged validity," *Journal of Behavioral Decision Making*, vol. 2, p. 81–94, 1989.
[33] I. Begg, A. P. Anas, and S. Farinacci, "Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth," *Journal of Experimental Psychology General*, vol. 121, no. 4, pp. 446-458, 1992.
[34] W.-C. Wang, N. Brashier, E. A. Wing, E. J. Marsh, and R. Cabeza, "On known unknowns: fluency and the neural mechanism of illusory truth," *Journal of Cognitive Neuroscience*, vol. 28, no. 5, pp. 739-746, 2016.
[35] L. K. Fazio, N. B. Brashier, K. B. Payne, and E. J. Marsh, "Knowledge does not protect against illusory truth," *Journal of Experimental Psychology*, vol. 144, no. 5, pp. 993-1002, 2015.
[36] E. Tulving, "Eposodic and semantic memory," in Organization of Memory, New York, Academic Press, 1972, pp. 381-402.
[37] C. M. Parks and J. P. Toth, "Fluency, familiarity, aging and the illusion of truth," *Neuropsychology*, pp. 225-253, 2006.
[38] M. S. McGlone and J. Tofighbakhsh, "Birds of a feather flock conjointly," *Psychological Science*, p. 424–428, 2000.
[39] R. Peterson, "A quantitative analysis of rating-scale response variability," *Marketing Letters*, vol. 8, pp. 9-21, 1997.
[40] Facebook, "Tips to Spot False News," 2017. [Online]. Available: https://web.facebook.com/help/188118808357379?_rdc=1&_rdr.