Fake Data and AI: Debunking Fake News to Educate and Enhance Media Literacy – A Study

Kavita Soni[(✉)] and Rekha Shelke

College of Journalism and Mass Communication, MGM University, Aurangabad, India
ksoni@mgmu.ac.in

Abstract. Everyone in this modern and technology-prone era relies on various online news sources for quick access of information. In addition, with the rise in popularity of social networking sites, within a short span of time, rumors circulate quickly among millions of users. Fake news is a threat to democratic societies and political systems by fostering hatred through a variety of methods, including satirical or fake data, imposter information, created content, fake links, false context, and manipulated content spread through social media platforms such as WhatsApp, Facebook, and Twitter [1] (Quandt, 2018). To halt the dissemination of fake content in emerging nations like India, it has become the need of the hour to educate every person about the debunking of false information, thus resulting in digital media literacy. The research paper is a study that is descriptive in nature, and it explores and analyses various available digital tools and technology for debunking virtual reality and fake news in the media. The study reveals that in India rise in machine based digital literacy is required to be familiar with reliable artificial intelligence supported fact-checking mechanisms to make people aware of the styles and techniques available for easily identifying and debunking fake news.

Keywords: Fake News · Debunking · Artificial Intelligence · Machine Learning

1 Introduction

Fake news has become increasingly popular on social media platforms in recent years. The spread of fake news has the intention of misleading targeted audiences, resulting in more and more dangerous consequences. It is getting harder to tell what data is reliable from what is deceptive on digital platforms because of the enormous growth of information resulting in the explosion of fake news. Popular social media platforms like Facebook and Twitter were purposely created to make it simple for users to publish any article, regardless of whether it is true or not, with the aim of spreading it widely [2]. Our society has been affected by disinformation, which is why studies on fake news identification are piquing scholars' attention around the world [3]. This research paper will mainly focus on the use of artificial intelligence for automatically debunking fake images, thus aiming to educate and enhance media literacy.

2 Fake News

The intentional spread of fake content information via social media platform is known as fake news. False information spreads like wildfire. The fact that if a false news website is taken down, another one will appear right afterwards to fill the gap [4]. Furthermore, because it spreads so quickly, fake news can merge with legitimate information and become unidentifiable. People can download content from websites, share it with others, then re-share it too, and by the time the day is up, the fake information will have spread so far from the original source that it will be impossible to differentiate it from the actual news [5]. The nature of social media differs significantly compared to that of traditional media. Although users are not responsible for the colourful and engaging content but encourages users to reproduce digital content for distribution. As a result, it is still doubtful whether social media platforms are real [6]. On the other hand, consumers of traditional media have little influence over how the content is distributed. Actually, the authenticity and truthfulness of a report or news article is the duty of the editors of newspapers and electronic media [7]. One of the basic bases of journalism includes the verification of the truth. Unfortunately, nobody is accountable for the post on social media, and it is challenging to identify who created the fake content [8]. The seriousness of the rise of fake news in the global south was highlighted by Eric Wishart, Editor of AFP. WhatsApp and other social media messaging apps disseminate misleading info and fake news, threatening the credibility and trust of journalists [9].

According to the Cambridge Dictionary, Fact-checking is the practise of ensuring that all facts related to a specific event, including images, videos, or graphics, are correct [10]. To "debunk" something is to expose fake news or show that something is less critical, more beneficial, or true than it has been portrayed as being. Fact-checking and debunking are frequently seen as an essential, if ineffective, response in a time when "fake news" is widely disseminated.

Humans are likely to be biased and poor fake news detectors [11]. People tend to believe information that does not contradict their beliefs, and they will share information just because a celebrity tweeted it (or by a celebrity fake account). It is inadequate, however, to rely solely on automated detection techniques or even artificial intelligence. Based on patterns in training data, machine learning in AI identifies whether content is accurate or false, but it cannot tell whether a statement is a joke or the truth. Likewise, it might be challenging to debunk fake news using non-textual resources like videos, images, and graphs. Mykhailo Granik implemented a fake news detection system using a Naive Bayes classifier in their paper. The system was tested using a dataset of Facebook news posts, and it achieved a classification accuracy of 74% [12].

3 Research Methodology

This research paper is an in-depth analysis of techniques used for debunking fake news. Hence, qualitative methodology was used. In addition, a review of related findings, threats, trends, and digital initiatives to combat fake news was carried. A descriptive and qualitative analysis was initiated as a part of study of various tools used in debunking fake news. This paper presents comprehensive details of fake news detection techniques

along with the identification of available data sets. Study will discuss the various artificial intelligence methods along with their comparative study and research challenges and potential directions for the detection of fake news.

4 AI and Debunking Fake News

Artificial intelligence can be used to produce deep fakes of audio, visual, and video content that compel individuals to express themselves and act in ways they would not have done otherwise. Four critical competencies are typically needed to debunk fake news: reverse image search, smart keyword search, excellent powers of observation, and fundamental journalism competencies. Before verification of any information received, one has to understand, think, and question himself about the authenticity of the information received. Information obtained from Twitter or Facebook can be verified by looking up a variety of information about its followers, such as the person who is following, the legitimacy of the account, history, and connections to other individuals and websites. Some websites, such as www.twitonomy, www.twitteraudit.com, and www.foller.me, assist the user in initially determining the reliability of information. For verification of photos, Google reverse image plugins can be helpful. Websites like www.tineye do the same thing with added filters. Yandex plugin is useful for identifying photoshopped images. While verifying photos, consider the scale and brightness in the image, such as whether it is falling on everyone and everything is in the same way or if there is a difference, as well as the details of weather, equipment, proportions, and so on may also help to analyze the photo. After verifying the information, the next step would be to search and check the credibility of the source. This observation involves knowing specific details about the original uploader's holding. Sites such as whopostedwhat.com, Google search, graph.tips, foller.me, and twitonomy can be useful in learning more about the uploader's position, reason, motives etc.

Date Verification is the next step for in-depth verification of images. With the help of Right-click on the photo basic information about the properties of images are known through which details of date of image created, modified and accessed is known. For referring the image to the Exif Viewer plugin on Chrome for the properties of the image websites like www.regex.info/exif.cgi (Jeffrey's image metadata viewer) is useful. Date of You Tube can be recognized with the help of date stamp as originally they are created in local time. Hence with the support of websites like www.amnestyusa.org/citizenevidence accurate date of videos can be identified. To check the weather in the image, assistance of websites like www.wolframalpha.comcan be benefitted.

Besides that, visual cues such as signs, clothing, landscape, structures, weather, logos, and vehicles are required for location verification. To cross-check visuals and location, use of Google Maps, Google Street View, or Wikimapia is preferred. Websites like www.firstdraftnews.com, www.bellingcat.com are excellent resources for learning about citizen journalism verification (Fig. 1).
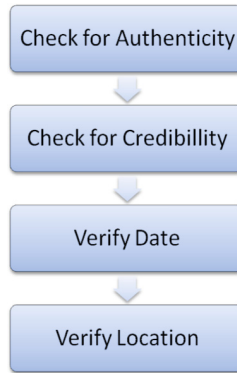
Fig. 1. Process of Debunking Fake News

5   AI Based Verification Tools

There are many online tools available to enhance these skills, like Invid/WeVerify, Google image search, TinEye search, Cydral search, Yandex search, Baidu, etc. Following are the examples of some verification tools with their details (Fig. 2).

6   Invid/We Verify Plugin

In 2021, the Invid/WeVerify plugin took first place in the US Department of Global Engagement Center's Paris Tech Challenge (GEC). According to the Poynter Institute, which hosts the International Fact-Checking Network, it is "one of the most powerful tools for spotting misinformation online" (IFCN). Invid/WeVerify is a verification "Swiss army knife," assisting journalists, fact-checkers, and human rights defenders in saving time and being more efficient in their fact-checking and debunking tasks on social media, particularly when verifying videos and images. Invid/WeVerify plugin has the following features:

1. Contextual information can be obtained from Facebook and YouTube videos.
2. Perform reverse image searches on Google, Bing, Baidu, Yandex, TinEye, and Reddit.
3. Video fragments from various social networks, including Facebook, Instagram, YouTube, Twitter, and Dailymotion.
4. Using a magnifying lens, you can improve and explore keyframes and images.
5. Read the metadata of videos and images.
6. Examine the video's copyrights.
7. Filter images with forensic filters.
8. Utilize timed intervals and multiple filters to query Twitter more efficiently (Fig. 3).

Across 202 countries more than 57,000 users use this plugin for debunking fake information. There are many advanced fact-checking AI based techniques available with the InVID verification plugin whose details are as follows.
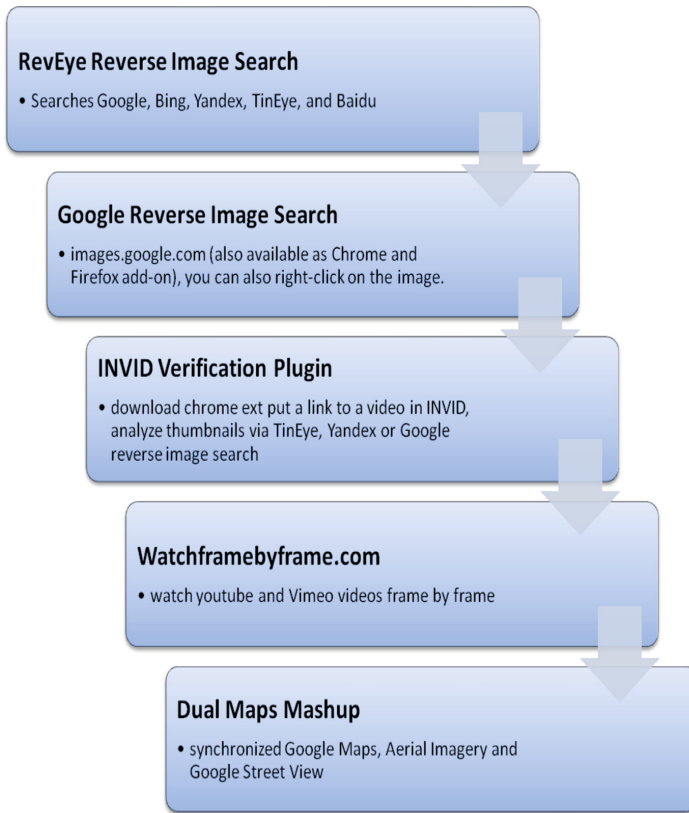
RevEye Reverse Image Search
- Searches Google, Bing, Yandex, TinEye, and Baidu

Google Reverse Image Search
- images.google.com (also available as Chrome and Firefox add-on), you can also right-click on the image.

INVID Verification Plugin
- download chrome ext put a link to a video in INVID, analyze thumbnails via TinEye, Yandex or Google reverse image search

Watchframebyframe.com
- watch youtube and Vimeo videos frame by frame

Dual Maps Mashup
- synchronized Google Maps, Aerial Imagery and Google Street View

Fig. 2. Various photo Image verification tools
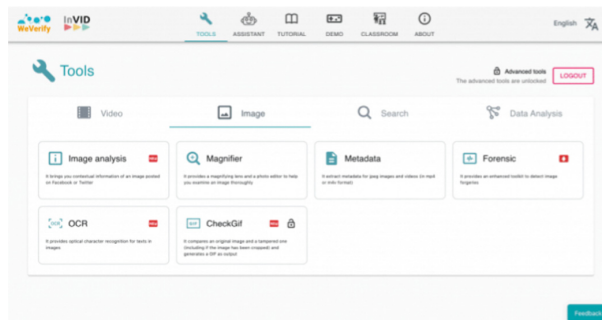
Fig. 3. InVID/WeVerify plugin

6.1 Assistant Techniques

It helps to analyze a webpage, an image or a video file and suggest which tool is useful in each case (Fig. 4).
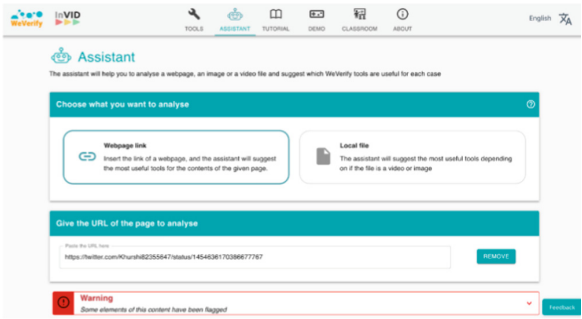
Fig. 4. Assistant Techniques
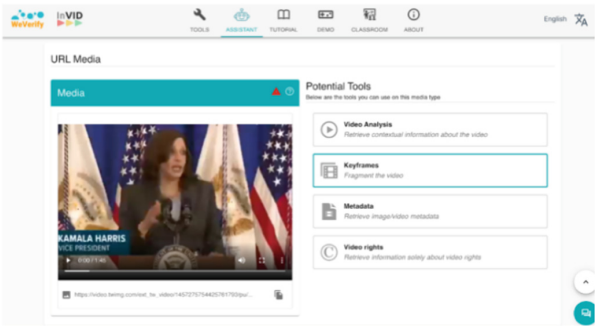
Fig. 5. Video Tool

6.2 Video Tools

It helps to evaluate a video in various forms like video analysis for contextual information, Keyframes for each fragment of video, Metadata for retrieving image or video metadata and video rights for recognizing the information solely about video rights (Fig. 5).

6.3 Image Tools

It helps to examine an image in various forms like image analysis for contextual information, magnifier for observing the image clearly, Forensic for detecting any image manipulation and Optical Character Recognition in Images (OCR) for extracting the text from the image (Fig. 6).

6.4 NLP Extraction Tool:

These tool attempts to build technique that know and take action to text or voice data—and respond with text or speech of their own—in much the same way humans do (Fig. 7).
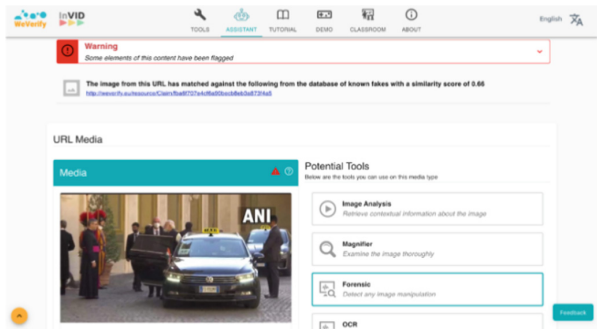
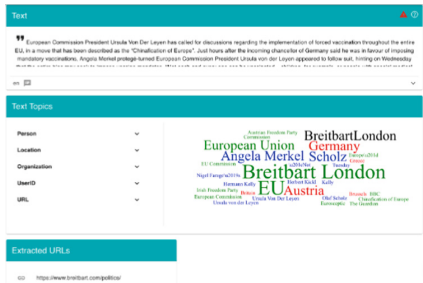**Fig. 6.** Image Tool
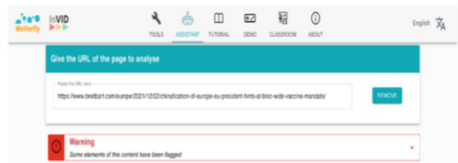
Fig. 7. NLP Extraction Tool

Fig. 8. Content Extraction Tool

6.5 Content Extraction Tool

It helps to analyze the content include in the piece of information received and suggests the related link from where the original information originates (Fig. 8).

6.6 Database of Known Fakes Extraction Tool

This tool guide users with various tools and offers a variety of other services such as link extraction, content matching against a credibility service, or the Database of Known Fakes (DBKF). The plugin contextual menu also allows you to search for similarity in the DBKF. It helps to extract tweet text as well as image or video links, recommend the most appropriate tool from the Verification plug-in toolset to the user, submit queries
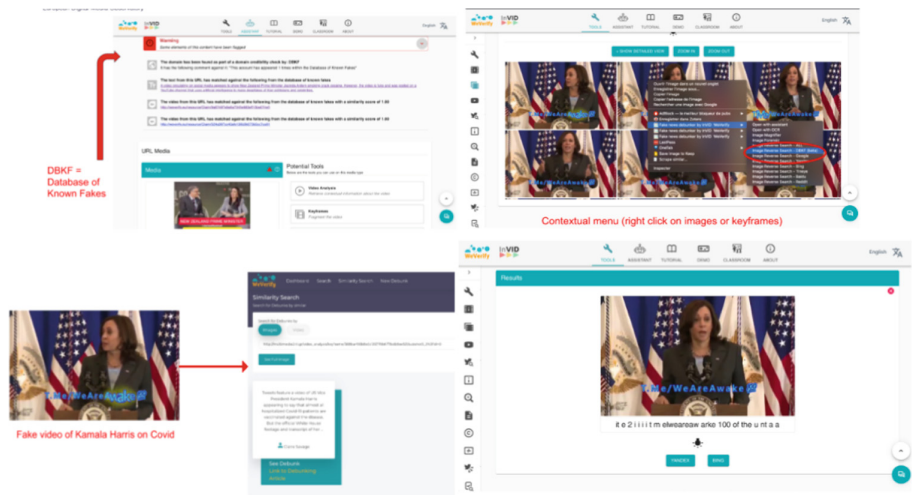
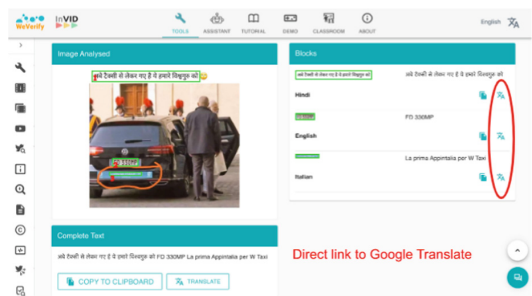**Fig. 9.** Database of known Fakes Extraction tool

Fig. 10. Multi-script OCR tool

to the database of previously debunked fakes in order to retrieve similar debunking and present them to the user (Fig. 9).

6.7 Multi-script OCR Tool

A web service for performing OCR specifically tuned to extracting the text from memes, adverts, and other images. It helps to convert scanned documents, such as images, into searchable and editable content (Fig. 10).

6.8 Forensic Analysis Tool

The enhanced forensic analysis toolkit aims to help in detecting alterations in manipulated images. Various enhancers and filters are available to highlight the particular suspicious area (Fig. 11).
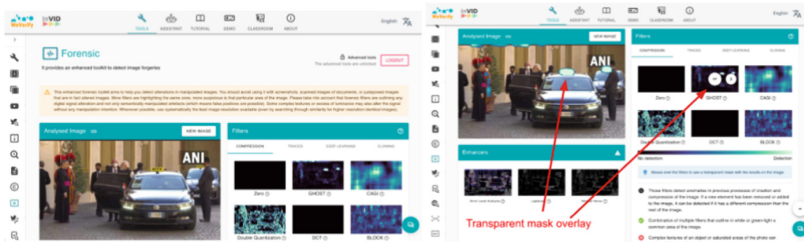
Fig. 11. Forensic Analysis tool

6.9 Advance Tool

The advance tools specially developed and reserved for journalist, fact checkers and researchers includes Twitter SNA (Social Network Analysis) and CrowdTangle SNA (Social Network Analysis). The Twitter SNA tool helps to search the element whether it is keyword, URL or hashtag along with the dates of twitted threads and their specific time zone. Various Index indicators like number of tweets, words of tweets, propagation timeline, retweeted users, most liked, active and mentioned users, bubble chart, heat map, most associated hashtags, socio-semantic graph, interaction graph and URL in tweets also help in verification of the information within the tweet. The Twitter SNA feature conducts a comprehensive analysis of disinformation on Twitter.

CSV Analysis can perform social network analysis from a CrowdTangle CSV export which includes details from Facebook and Instagram to help follow, analyze, and report on what's happening across social media. Various index indicators like metrics, propagation timeline, retweeted users, and most liked, active and mentioned users, heat map of tweets distribution, most associated hashtags, socio-semantic graph, interaction graph and URL also help in verification of the untruth information (Fig. 12).

6.10 Forensic Gif Export and Check Tool

It is an advanced feature available only to registered fact-checkers, journalists, and researchers that allows the creation of a GIF between a manipulated image and an original image in order to better reveal the manipulation (Fig. 13).
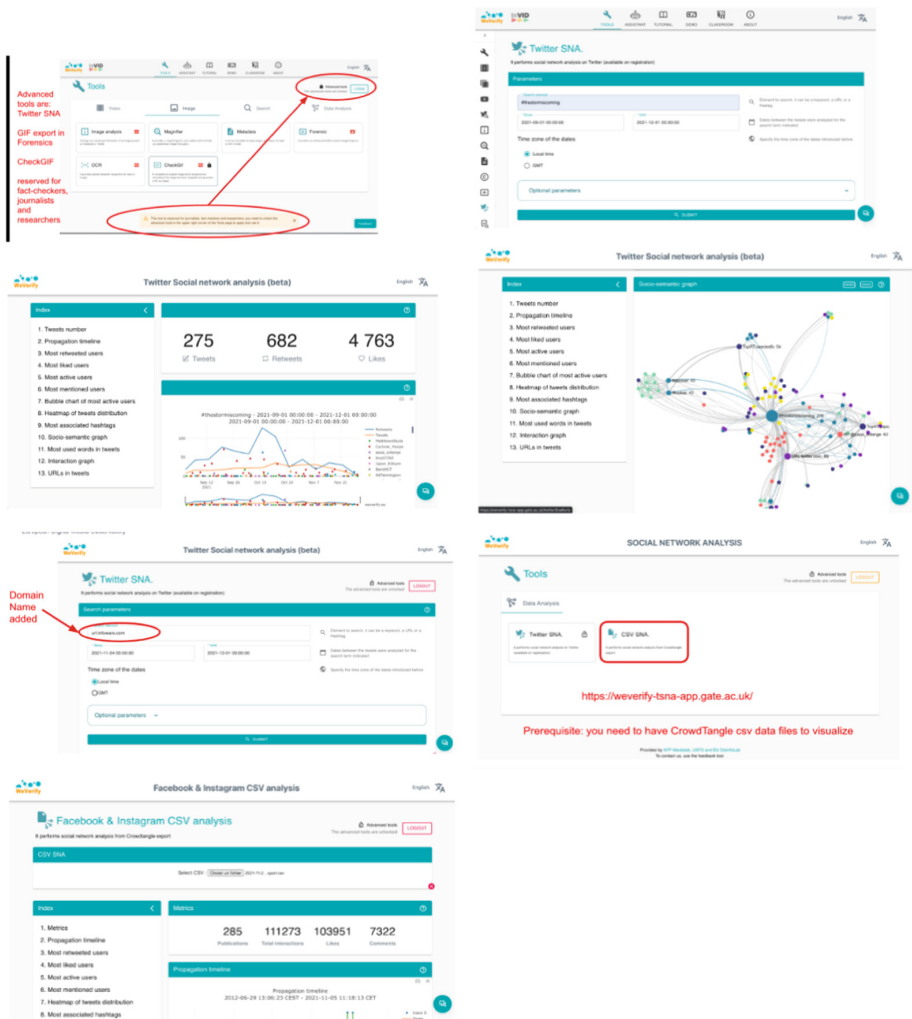
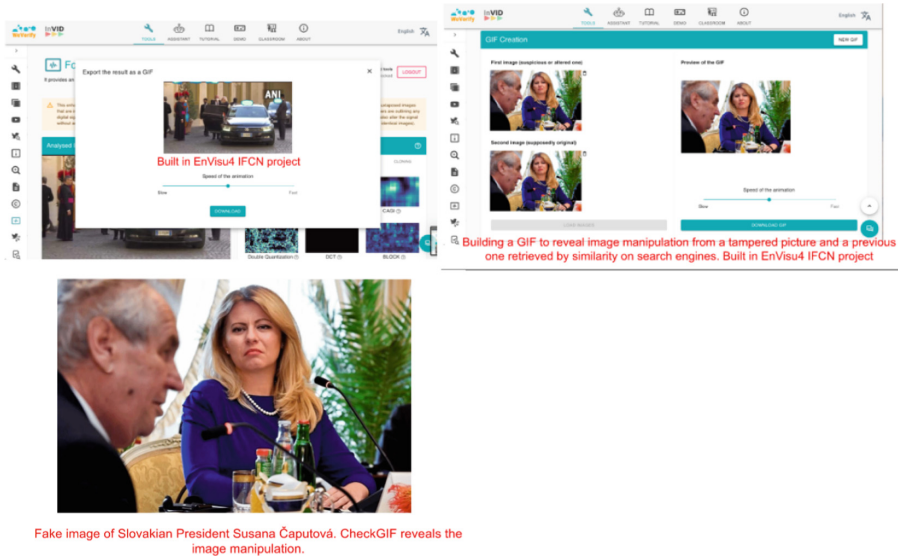**Fig. 12.** Advance tool (Twitter SNA and CSV Analysis)

Fig. 13. Forensic Gif Export and Check tool

7   Conclusion

The AI advancement has enabled the art of debunking fake data spread across internet. The technique of fake news detection on social media is very unique, and research is now being conducted in the hope that layman may discover more precise techniques to identify incorrect material in this expanding, fake-news-infested field. The study could help other researchers determine the best combination of techniques for effective identification of fake news on social media.

Debunking fake news can and will assist the public in making informed decisions, but it is not a stand-alone solution to the problem of combating disinformation. Just like human nature, understanding fake news is challenging. Therefore, the debunking of fake news necessitates multiple approaches. Assuming that the technology used to prevent fake news is the same technology used to create it, i.e. "AI." The study concludes that the masses should be made acquainted with the AI and machine learning tools to avoid falling prey to disinformation, misinformation and fake news.

References

1. Quandt, T. (2018, November 8). Dark Participation | Article | Media and Communication. Retrieved December 17, 2022, from https://www.cogitatiopress.com/mediaandcommunication/article/view/1519
2. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

3. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017, August 7). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved December 17, 2022, from https://arxiv.org/abs/1708.01967v3

4. Stahl, K. (2018, April 20). Fake news detection in social media. Retrieved December 17, 2022, from https://www.semanticscholar.org/paper/Fake-news-detection-in-social-media-Stahl/b202b4b7124vDpGt8SeHRSRhx2XpnAQPTu21HhXFB5

5. Rubin, V., & Chen, Y. (2016). *The UNDARK Magazine.* Retrieved from https://www.researchgate.net/publication/311065367_Education_and_Automation_Tools_for_Navigating_a_Sea_of_Fake_News

6. Kanozia, Dr. R. (2019). Analysis of Digital Tools and Technologies for Debunking Fake News. *Journal of Content, Community & Communication.* https://doi.org/10.31620/JCCC.06.19/16

7. Koohikamali, M., & Sidorova, A. (2017). Information Re-Sharing on Social Network Sites in the Age of Fake News. *Informing Science: The International Journal of an Emerging Transdiscipline*, *20*, 215–235. https://doi.org/10.28945/3871

8. Matt Carlson, N. U., Usher, N., & Carlson, M. (2018, November 8). The Midlife Crisis of the Network Society | Commentary | Media and Communication. Retrieved December 17, 2022, from https://www.cogitatiopress.com/mediaandcommunication/article/view/1751

9. Kvasnička, T. (2014). Asociaceaterotrombózy a trombofilie - genetickéaspekty. VnitrniLekarstvi, 60 (10), 880–884

10. fact-checking. (2022, December 14). Retrieved December 17, 2022, from https://dictionary.cambridge.org/dictionary/english/fact-checking

11. Salazar, A. (2020). *AI Tools on Fake News Detection: An Overview and Comparative Study.* Retrieved from https://www.researchgate.net/publication/343859545_AI_Tools_on_Fake_News_Detection_An_Overview_and_Comparative_Study

12. Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).* https://doi.org/10.1109/ukrcon.2017.8100379

13. Pamment, J., & Lindwall, A. (n.d.). FACT-CHECKING AND DEBUNKING A BEST PRACTICE GUIDE TO DEALING WITH DISINFORMATION. *NATO Strategic Communications Centre of Excellence.* NATO Strategic Communications Centre of Excellence. Retrieved from https://stratcomcoe.org/cuploads/pfiles/nato_stratcom_coe_fact-checking_and_debunking_02-02-2021-1.pdf

14. Semantic Scholar | AI-Powered Research Tool. (n.d.). Retrieved December 17, 2022, from https://www.semanticscholar.org/

15. https://downloads.bbc.co.uk/academy/academyfiles/Verification_links_handout.pdf

16. https://chrome.google.com/webstore/detail/fake-news-debunker-by-inv/mhccpoafgdgbhnjfhkcmgknndkeenfhe?hl=en

17. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news

18. https://weverify.eu/verification-plugin/