Identifying False Information in the Era of Big Data

Yikun Li[1,a,*]

[1]*Pioneer College Inner Mongolia University, Hohhot City, China*
a. 1912231134@mail.sit.edu.cn
*corresponding author

Abstract: The emergence and dissemination of false information in the era of big data has greatly affected people's lives, and identifying and correcting such false information is critical to the safety of personal information property and public health. To understand how the public is "influenced" by fake news, this paper surveyed 303 respondents from different ages. Controlled experiments were conducted based on which social platforms they were active on a daily basis, whether the information they often received could be identified as true or false, and whether they were willing to share this information with those around them. In each experiment, the study predicted the accuracy of different groups' judgments of true or false messages, as well as their willingness to share the messages with those around them. In both experiments, information from authoritative organizations or institutions significantly increased the accuracy of participants' judgments and their willingness to share the information with others. In addition, the results of the post-experiment questionnaire showed that participants would consider the authenticity and reliability of the corrected information sources and would confirm them by themselves.

Keywords: false Information, big data, public health

1.   Introduction

With the development of the global digital age, especially the development of the Internet. Information has become commonplace in the public's daily life, and the public has more and more channels to find the information they want, and to get it in an efficient and fast way. However, they do not screen the information to see if the source is reliable or if it comes from an authoritative publisher or government agency. In such cases, people are vulnerable to deception and can lose their information or money.

How to identify false information and at the same time how to reduce the appearance of such information has always been a problem of social research. It is known that there is a correlation between disinformation and human emotions, and the higher the anxiety level of a person, the stronger the acceptance of disinformation. In addition, people are more sensitive to fake news after correction by authorities [1]. Disinformation may cause an event or thing to be misunderstood by the public or society so that things do not happen in the direction one would expect, or worse, have the opposite effect and prevent the event from happening [2]. How to stop the spread of disinformation and how to keep it from being adopted by the public or identify it as misinformation and reject it is an extremely important key. For example, by setting up a problem to identify misinformation adoption or rejection stance classification, a neural language processing system was designed to operate on

social platforms using graph attention networks relying on lexical, affective and semantic knowledge to stop the spread of disinformation [3]. Also depending on the living environment of the public at the time of the new crown epidemic, the presence or absence of others had an effect on their mood and the degree of disinformation belief. For example, people who are alone at home and depressed are more likely to receive disinformation, while multiple people together are less likely to be affected. The middle-aged and elderly people, because of their aging cognition, preference for positive information and lack of precision in processing detailed information, will have more obvious characteristics towards false information or advertisements, such as choosing to accept false information in front of ancient remedies and cures or health products. Also, the stance of whether false information can be automatically classified in social platforms is the solution [4]. And at the same time, self-association tends to make the public interpret certain information more incorrectly, resulting in the dissemination of information that is supposed to occur truthfully into false information and unconsciously without the knowledge of that communicator [5]. So, it is important process to help people try to understand the cognitive process and increase the sensitivity of people to remember and believe misinformation [6-10]

Therefore, this paper will conduct a longitudinal study to fill the potential research gap in the current academia. Firstly, A questionnaire with 22 questions was designed. This study convened 303 survey respondents. The study focused on how people socialize on media platforms on a daily basis and conducted a series of investigations on the content of disinformation. The study focused on how people socialize on media platforms on a daily basis and conducted a series of investigations on the content of disinformation. The main questions were whether the individuals were able to determine the areas in which false information existed and asked about the false information they encountered in Covid-19, how they dealt with it when they encountered it and in what way it mostly appeared, and finally how the individual's obtained clarification of the false information and what they thought were the main factors that interfered with them. And these respondents were divided into two groups, in the first group the questionnaire was marked with the introduction of false information and how to identify false information, and at the end of the questionnaire several news were set up with real information and false information for the respondents in this version of the questionnaire to judge the answers, in the second book of the questionnaire, the content marked with the introduction, false information and how to identify false information was deleted in After the normal completion of the questionnaire, the same let the investigators to make judgments of real news and false news. The two versions of the questionnaire were analyzed and compared to determine whether the public could improve their sensitivity to false information after correction.

2.   Questionnaire Design

The questionnaire for this experiment was filled out by calling the respondents online through the Internet, so the respondents were not limited to one place, but different groups from different regions and different age groups across the country to fill out the questionnaire, so as to ensure the scope, authenticity and reliability of the experimental data. The difference between questionnaire 1 and questionnaire 2 is that the concept of how to identify false information and disinformation was added to questionnaire 1, while it was not reflected in questionnaire 2. In this study, a table was used to show more clearly the similarities and differences between the two versions of the questionnaire, where a check mark indicates that the content is present in both, while a wrong mark indicates that it is not present (See Table 1).

Table 1: Summary of the questions.

Questions	Questionnaire 1	Questionnaire 2
The questionnaire was marked with the introduction of false i nformation and how to identify false information	√	×
Age distribution of residence	√	√
2main sources of information	√	√
3common social media platfor ms' time of day	√	√
4whether individuals were will ing to share information	√	√
5whether individuals judged w hether false information existe d during the sharing process	√	√
6 whether the individuals were able to determine the areas in which false information existe d	√	√
7asked about the false informa tion they encountered in Covid -19	√	√
8how they dealt with it when t hey encountered it	√	√
9in what way it mostly appeare d	√	√
10how the individuals obtaine d clarification of the false infor mation	√	√
11what they thought were the main factors that interfered wit h them	√	√

3. Questionnaire Results

After the recall, it was found that the respondents of questionnaire 1 were significantly more sensitive to false information and correct after correction than questionnaire 2. Similarly, the difference between the data of the two responses in the previous questions is not significant, so it can be proved that the group that was corrected and is able to improve their sensitivity to false information, while the group that was not corrected for the word for the judgment of false information, the degree of association is much higher than the former (See Table 2).

Table 2: Results of the survey (1).

question	questionnaire 1	data	questionnaire 2	data
Listening to senior health seminars and purchasing follow-up medical products can help seniors' mental health and physical health	√	False 89.76% True 10.24%	√	False 67.15% True 32.85%
"Feed added western medicine, can make the animals raised silly eat sound sleep fierce growth, but the antibiotic in the meat has residual, people eat with antibiotics after the meat, or produce "drug resistance". In the long run, it may make some kind of germs, viruses to produce resistance, which will lead to the whole human can no longer effectively resist disease."	√	False 81.02% True 18.98%	√	False 55.42% True 44.58%
"Vaccination reduces rates of serious illness and mortality in older adults."	√	True 62.05% False 37.95%	√	False 42.34% True 57.66%
"Wearing two masks at the same time can provide double the protection."	√	False 72.89% True 27.11%	√	False 67.15% True 32.85%

In the survey, it can be found that most people still choose to spend time with family and friends, only a small number of people choose to spend time alone, while there is no difference between the answers of the two questionnaires in the most frequently used social media platforms, except for a very small number of traditional media, such as newspapers and radio, more and more people choose digital media. Secondly in terms of time spent using social media each time period is generally present between two and four hours. Moreover, people are very willing to share the advertisements or information they see that they are interested in with their friends and family members, and they also

think that there is false information in the process of exposure to information in general. According to the data, the respondents believe that there is false information in every field, and in the era of the new crown epidemic, the respondents are misled by information on the Internet or by people around them, and some of them even choose to believe false information. When they encounter false information, most people choose to warn their neighbors and family members to be careful of such information, while some choose to stick to their own judgment, and some choose to ignore it. At the same time, according to the data, false information mostly relies on social media and appears in various forms, including but not limited to pictures, videos, texts and links. Finally, in the process of spreading false information, credible institutions or governments as well as friends or family members around individuals act as information clarifiers (See Table 3).

Table 3: Results of the survey (2).

Question	question naire 1	data	question naire 2	data
Your current living situation	√	Living alone at home 20.48% Living with parents/ partners 79.52%	√	Living alone at home 5.84% Living with parents / partners 94.16%
What is the main source of information you usually obtain?	√	Social media 66.87% Television/newspaper / broadcasting 33.13%	√	Social media 74.45% Television/newspaper /broadcasting 25.55%
How long does your daily media usage last?	√	0.5h<duration<2h 44.58% 2h<duration<4h 44.58%	√	0.5h<duration<2h 55.47% 2h<duration<4h 60.58%
When you see advertisements or information that you are interested in, are you willing to share it with your family and friends around you?	√	Frequent/Occasional 90.96% little or no 9.04%	√	Frequent/Occasional 89.05% little or no 10.95%
Do you think there is any false information in the information that is usually accessed through the media?	√	There are some/many / A lot 80.12% No/very little 19.88%	√	There are some /many/A lot 56.93% No/very little 43.07%

Table 3: (continued).

What specific areas do you think false information on the internet exists in?	√	A. Medical categories such as miraculous doctors treating diseases and selling ancient medicinal herbs 52.41% B. Shopping categories such as trial sales of high-end health products 46.99% C. Free, low-priced travel and other travel activities 50.6% D. High interest funds, investments, loans, and other financial categories 51.2% E. Counterfeit official, false advertising and other Phone fraud 51.2% F. Social, public topics and other false news 52.41%	√	A. Medical categories such as miraculous doctors treating diseases and selling ancient medicinal herbs 44.53%% B. Shopping categories such as trial sales of high-end health products 43.07% C. Free, low-priced travel and other travel activities 42.34% D. High interest funds, investments, loans, and other financial categories 35.04% E. Counterfeit official, false advertising and other Phone fraud 35.77% F. Social, public topics and other false news 40.15%
In the Covid-19 era, did you receive misleading information from the internet or people around you? (For example, xx medicine can treat or alleviate symptoms)	√	Yes, and take action to purchase 12.05% Yes, but not adopted 64.46% Yes, but personally verify and inform the message sharers 22.89% No 6.02%	√	Yes, and take action to purchase 67.88% Yes, but not adopted 64.23% Yes, but personally verify and inform the message sharers 67.88% No 0%
What is your usual handling method when encountering information that has been proven to be false?	√	Ignore and stand idly by 30.12% Tell family and friends not to trust others 38.55% Stick to your own judgment and continue forwarding 16.27% Reminder for forwarding information, reporting this message 27.71%	√	Ignore and stand idly by 45.26% Tell family and friends not to trust others 48.91% Stick to your own judgment and continue forwarding 51.82% Reminder for forwarding information, reporting this message 46.72%%

Table 3: (continued).

What is the form in which most of the false information you are exposed is presented	√	Picture 46.99% Short video 52.41% Medium length video 37.95% Text 38.55% Popup 45.78% Links on Moments 49.4%	√	Picture 55.47% Short video 38.69% Medium length video 43.8% Text 43.07% Popup 42.34% Links on Moments 40.88%
What methods (platforms) do you usually use to clarify false information?	√	Communication among classmates, colleagues and friends 60.84% Communication between children, parents, and other family members 40.96% Communication between communities and other social groups 43.37% Paper tools such as newspapers and magazines 35.54% Electronic tools such as radio and television 49.4% WeChat, Weibo, Today's Release and other online tools 54.82% Other 2.41%	√	Communication among classmates, colleagues and friends 51.82% Communication between children, parents, and other family members 48.18% Communication between communities and other social groups 53.28% Paper tools such as newspapers and magazines 45.99% Electronic tools such as radio and television 49.64% WeChat, Weibo, Today's Release and other online tools 46.72% Other 0%

4. Discussion

According to the data of questionnaire 1 and questionnaire 2, we found that the ability of the experimental subjects to identify false information was significantly higher in one than in two, and this may be related to the fact that the credible institutions or governments around these questionnaire interviewers and the friends or family members around the individuals played a big role in clarifying the information. It may also relate to the automatic identification of the platform's algorithm in the social media used by the respondents and the self-censorship of some hashtags, as well as to the behavioral correction of the subjects at the beginning of the questionnaire and the continuous input of false information content and concepts to them during the survey.

5.   Conclusion

With globalization and the development of the Internet, disinformation has reached a point in its course where we cannot ignore the problem. With the end of the new crown epidemic, we have a lot of evidence that disinformation affects not only people's daily lives, but also their emotions, judgments, and the development of the real world. In this study, we investigated whether individuals could detect and avoid the hidden risks of disinformation when they are corrected by people around them or by a credible institution or government. Questionnaires were completed by a large number of respondents, and experiments were conducted. Questionnaire 1 included information from authorities or government on how to distinguish and identify false information, while questionnaire 2 did not include such information. The final result was that after correction, the respondents did increase their sensitivity to disinformation and achieved the expected results of the experiment. The error rate of the respondents in questionnaire 2 also met the experiment's expectation without reminders and corrections.

There may be some shortcomings in the study. First, the number of questionnaires in this study is small. Second, the scope of the sample collection may not be scientific. Third, the questionnaire may be unscientific. The deficiencies deserve further investigations.

References

[1]   Seo, H., Xiong, A., Lee, S., and Lee, D. (2022) If You Have a Reliable Source, Say Something: Effects of Correction Comments on COVID-19 Misinformation. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 896-907.

[2]   Sharma, K., Zhang, Y., and Liu, Y. (2022) COVID-19 Vaccine Misinformation Campaigns and Social Media Narratives. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 920-931.

[3]   Weinzierl, M., Hopfer, S., and Harabagiu, S. M. (2021) Misinformation Adoption or Rejection in the Era of COVID-19. Proceedings of the International AAAI Conference on Web and Social Media, 15(1), 787-795.

[4]   Zeng, L., Starbird, K., and Spiro, E. (2021) Unconfirmed: Classifying Rumor Stance in Crisis-Related Social Media Messages. Proceedings of the International AAAI Conference on Web and Social Media, 10(1), 747-750.

[5]   Lee, S., Seo, H., Lee, D., and Xiong, A. (2023) Associative Inference Can Increase People's Susceptibility to Misinformation. Proceedings of the International AAAI Conference on Web and Social Media, 17(1), 530-541.

[6]   Ling, C., Gummadi, K. P., and Zannettou, S. (2023) "Learn the Facts about COVID-19": Analyzing the Use of Warning Labels on TikTok Videos. Proceedings of the International AAAI Conference on Web and Social Media, 17(1), 554-565.

[7]   Papadamou, K., Zannettou, S., Blackburn, J., Cristofaro, E. D., Stringhini, G., and Sirivianos, M. (2022) "It Is Just a Flu": Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 723-734.

[8]   Abdali, S., Gurav, R., Menon, S., Fonseca, D., Entezari, N., Shah, N., and Papalexakis, E. E. (2021) Identifying Misinformation from Website Screenshots. Proceedings of the International AAAI Conference on Web and Social Media, 15(1), 2-13.

[9]   Lu, C., Wang, B. H., Tian, X., et al. (2022) The effects of closure state and emotion on the degree of belief of online rumors with different potencies during the New Crown epidemic. Chinese Psychological Association.

[10] Xing, T. (2020) The effects of false advertising and product type on the persuasion effect of the elderly. Thesis for master's degree, Tianjin Normal University.