Approaches to Identify Vulnerabilities to Misinformation: A Research Agenda

NATTAPAT BOONPRAKONG, University of Melbourne, Australia
BENJAMIN TAG, University of Melbourne, Australia
TILMAN DINGLER, University of Melbourne, Australia

Given the prevalence of online misinformation and our scarce cognitive capacity, Internet users have been shown to frequently fall victim to such information. As some studies have investigated psychological factors that make people susceptible to believe or share misinformation, some ongoing research further put these findings into practice by objectively identifying when and which users are vulnerable to misinformation. In this position paper, we highlight two ongoing avenues of research to identify vulnerable users: detecting cognitive biases and exploring misinformation spreaders. We also discuss the potential implications of these objective approaches: discovering more cohorts of vulnerable users and prompting interventions to more effectively address the right group of users. Lastly, we point out two of the understudied contexts for misinformation vulnerability research as opportunities for future research.

CCS Concepts:
• Human-centered computing → Human computer interaction (HCI).

Additional Key Words and Phrases: misinformation; vulnerable users; user profiling; cognitive bias

ACM Reference Format:
Nattapat Boonprakong, Benjamin Tag, and Tilman Dingler. 2022. Approaches to Identify Vulnerabilities to Misinformation: A Research Agenda. In CHI ’22: ACM CHI Conference on Human Factors in Computing Systems, April 30–May 6, 2022, New Orleans, LA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

There is a sheer amount of misinformation circulating on the Internet nowadays. Defined as false or misleading information, misinformation misinforms people and bears many of the harmful effects to individuals and society [19]. More recently, misinformation has undermined public health guidance about COVID-19 [31]. Through Internet users who consume online media, misinformation has been shown to spread faster than factually accurate information [38]. This phenomenon, where users believe and share misinformation without taking deliberate considerations, can be attributed to not only the presence of misinformation but also the psychological vulnerabilities of information consumers [9]. Humans are not always rational [35]; therefore, they are prone to make erroneous decisions, like misinterpreting news or sharing unverified information.

There are two main actions regarding misinformation: believing and sharing. In the literature, research has explored several user-side factors that govern the susceptibility of people to believe online misinformation. Cognitive biases—mental shortcuts we use to achieve faster but less deliberate decisions—have been marked as factors that make people be gullible to false or unverified information [26]. Martel et al. [21] have shown that reliance on emotion indicates higher belief in false information. Some research have shown that political
partisans [25] or those who are older than 65 years [3] are prone to believing fake news. Roozenbeek et al. [31] found that higher numeracy skills and better trust in science are indicators of lower susceptibility to misinformation.

On the other hand, some research has investigated why people share unverified information online. Motivated by [8, 28], Karami et al. [17] listed five motivational factors in spreading fake news: uncertainty, anxiety, lack of control, relationship enhancement, and social rank. Laato et al. [18] found that trust in online information and information overload are strong indicators of sharing unverified information. Chen et al. [5] suggested that people share misinformation because of social factors, i.e., self-expression and socialization. In addition, Avram et al. [1] found that social engagement metrics (i.e., the number of likes or shares) increased the tendency to share misinformative contents.

Based on the above-mentioned psychological findings, some research have put these findings into practice by building computing systems that objectively identify potential misinformation spreaders based on their online and offline behaviours [13, 17, 36]. Unlike previous approaches that rely on self-report measures, these approaches quantify how people behave with information and produce insights about their specific behaviours regarding to misinformation. Through behavioural and psycho-physiological measures, these approaches can explicitly identify when and which users are prone to believe or spread misinformation.

Different individuals have different backgrounds and different levels of cognitive capacity. Therefore, some information consumers may be more prone than others to misinformation. Echoing the findings in Geeng et al. [12] that different design interventions may be effective for different users, the objective approaches to identify vulnerable users would also provide a guide on which misinformation intervention should be deployed to which group of users. Ultimately, interventions would be made more effective as they address the right group of users.

In this position paper, we discuss empirical approaches to identifying people who may be susceptible to misinformation: detecting cognitive biases and profiling vulnerable users. We also highlight their implications of identifying more cohorts of vulnerable users and prompting interventions to address the right group of users. Lastly, we elaborate on challenges and future avenues to investigate vulnerabilities to misinformation. We intend to generate a fruitful discussion in this workshop about the need to explore potential victims of misinformation, identify their vulnerabilities, and employ the right interventions to address the right user cohorts.

2 APPROACHES TO IDENTIFYING VULNERABILITIES TO MISINFORMATION

Various studies have investigated why people believe and share misinformation on the Internet. However, most of the studies addressed such questions based on self-report measures, e.g., questionnaires, which are prone to many limitations such as subjectiveness. On the other hand, some research has proposed objective approaches to identifying users who may be susceptible to misinformation. Combining prior psychological findings and behavioural measures, these research efforts explicitly identify which and when users tend to believe and share misinformation. In this paper, we discuss two promising approaches: detecting cognitive bias detection and profiling social media users.

2.1 Cognitive Bias Detection

Humans possess limited cognitive capacity and employ cognitive biases as mental shortcuts, which can lead to irrational judgements. One prominent example is “confirmation bias” (also known as selective exposure) as a tendency to seek solely information to support one’s perspectives or expectations, while ignoring other dissenting information [24]. Taking the advantage of social media platforms that inform users with contents that reinforce their viewpoints, confirmation bias is easily seen on social media platforms as many users tend to consume and spread content items that match their belief without checking its veracity [39].
Addressing cognitive biases might be a feasible approach to identify when the user is susceptible to misinformation. Methods to detect confirmation bias have traditionally relied on self-reports [6], for example, asking users about their news consumption habits [11]. Recent approaches, on the other hand, have used behavioral measures to objectively determine confirmation bias (e.g., time spent on reading news articles). In addition, some research employed psycho-physiological signals as objective measures of confirmation bias, like eye tracking [20, 36] and electroencephalography [22].

However, we note that research on cognitive bias detection has been limited. While we have seen some research quantifying confirmation bias, little attention has been paid to investigate the detection of other cognitive biases that also help the spread of misinformation. The bandwagon effect, for example, is a tendency to adapt certain behaviours (e.g., believing a news piece) because many other people are doing the same. There have been some research efforts in the context of information retrieval to detect such phenomena [14]; thus, it is promising to see the quantification of bandwagon effects in the realm of online information consumption.

2.2 Profiling Vulnerable Users

Some ongoing research focused on empirical, objective approaches to profiling potential fake news believers and spreaders by using data mining and machine learning algorithms on real-world social media data. Shu et al. [34] extracted social media user profiles into explicit features (metadata, for example, the number of posts and the number of followers) and implicit features (e.g., age, gender, personality) and used them to infer users who are prone to believing fake news. Giachanou et al. [13] combined linguistic and the Big Five Personality scores in users’ tweets to classify Twitter users who spread fake news. Karami et al. [17] developed metrics to measure how likely a Twitter user will spread false information. Informed by research in psychology [8, 28], they quantified each user profile based on the five categories of motivational features: uncertainty, emotions, lack of control, relationship enhancement, and social rank. Furthermore, this problem led to a collective effort in a challenge at the PAN at CLEF 2020 workshop1 where several machine learning approaches to detect fake news spreaders were proposed [29].

3 FUTURE IMPLICATIONS: DESIGNING FOR DIFFERENT VULNERABILITIES

In this section, we discuss future implications from each of the two presented approaches to identifying those who are vulnerable to misinformation.

3.1 Cognitive Bias Reflection

The ability to detect cognitive biases implies that we can prompt users to reflect on their susceptibility to misinformation. (or, in the case of confirmation bias, information that confirms their opinion) Furthermore, interventions could be employed to address users who are subject to cognitive biases. On the other hand, we would be able to more accurately place misinformation interventions on contents that induce users’ cognitive biases. For instance, interface designers may place a warning message that prompts users awareness (e.g., “This content might reinforce your belief” [30]) or a nudge (e.g., recommending alternative contents to balance out the viewpoints [37]) on online contents/posts/tweets that pose cognitive biases and susceptibility to believe and spread misinformation.

3.2 Data-driven Identification of Vulnerable Users

Online behaviour data provide opportunities to model and profile users who may be more susceptible to misinformation. Data-driven approaches could identify more features associated with the users’ tendency to spread

1https://pan.webis.de/clef20/pan20-web/
misinformation online. Linguistic features, for example, might be able to tell the user’s age cohort [23] or political orientation [27].

Furthermore, with the abundance of features, we could perform clustering on the feature vectors of user profiles to explore user groups that are represented by clusters of data. Based on the t-SNE visualization of user profiles in Karami et al. [17] using PolitiFact and GossipCop datasets [33] (see Figure 1), we may be able to imply that there are some fake news spreaders who exhibited some similarities with those who did not share fake news because their data points are closely located on the representation space. Further investigations would reveal the features of such user groups and proper intervention methods to address them.

The identified vulnerabilities would also inform the design of misinformation interventions that is personalised or targeted according to users’ behaviour on platforms. At the same time, misinformation interventions would be made more explainable to the users as interventions are attributed by the vulnerabilities informed by the empirical results.

4 CHALLENGES: UNDERSTUDIED CONTEXT

Although approaches exist to detect potential consumers and spreaders misinformation, there is more avenues for researchers to investigate vulnerabilities to misinformation. In this section, we discuss two challenges: visual-based media and instant messaging services. Visual information and instant messaging are understudied contexts in misinformation research. Therefore, we argue that more research is needed to address people’s vulnerabilities to misinformation in the forms of visuals and private messaging.

4.1 Visual-based Media

Visual information (i.e., images and videos) is one of the most prevalent format of misinformation [4] and it is more likely to persist in our memory [7]. However, there has been limited research on visual misinformation [32], including those in the context of vulnerabilities to believe and disseminate visual misinformation. In the following, we discuss challenges regarding to the two approaches presented in this paper.

In the realm of cognitive bias detection, some research investigated confirmation bias in viewing visual stimuli. Marquart et al. [20] examined users’ selective exposure to political advertisement posters via eye-tracking data. The paper revealed significant effects of political predisposition on time spent looking political ads. From this paper, we argue that gaze data are promising means to identify patterns of visual information consumption. Gaze
approaches to identify vulnerabilities to misinformation: a research agenda


...

Additionally, there has been little research that investigates potential spreaders of misinformation on visual-based platforms. Yet, Hussein et al. [16] investigated YouTube’s personalisation algorithm through search engine audits and found that, while demographics (i.e., age, gender, and location) did not affect the amount of misinformation presented, users with watch history on misinformative topics (e.g., 9/11 or flat earth conspiracies) tended to be exposed to more of such videos. We suggest that further investigation is needed to assess the vulnerabilities of conspiracy video consumers. We also argue that visual-based media pose some challenges for user profile investigation. One prominent challenge is that the meaning conveyed by visual information can be highly contextual [32]. For example, one might share a video containing false claims with accurate user corrections in a caption.

4.2 Instant Messaging

Instant messaging (IM) applications (for instance, WhatsApp, WeChat, and Telegram) are one of the most used services on mobile phones. However, such services are fertile grounds for the spread of misinformation and propaganda [2, 10]. The role of IM services on misinformation has been understudied and obtaining real-world data has been difficult due to its private, encrypted nature.

We argue that cognitive biases exist in the interactions between users and messages. Users may forward message contents that confirm their beliefs to other users or chat groups. At the same time, some users may belong in one or more chat groups, which consist of people of same interests, e.g., liberal or conservative politics. In such cases, private group chats may be amplification venues for contents that tell predominantly one side of every story. It is therefore interesting to see the detection of cognitive biases in IM settings and interventions that invite users to reflect themselves on, for example, their information diet.

While there has been some studies that explored motivations in sharing fake news on IM platforms [15], there is little or no research on approaches to objectively identify IM users who may be susceptible to misinformation. Although data on IM platforms are hardly accessible, it is encouraging to see many of the psychological aspects (for example, the pace of communication and emotion expressed as a group) be taken into account when designing systems that detect potential spreaders of misinformation.

5 CONCLUDING REMARKS

Misinformation does its work in deceiving and misinforming users. However, the question remains how people fall victim of misinformation. Various research efforts have explored psychological factors involved; furthermore, some ongoing research have brought these insights into practice. In this position paper, we discuss two promising approaches to identify people’s vulnerabilities towards believing and spreading misinformation. Through cognitive bias detection and user profiling, we explore their implications in (1) prompting misinformation interventions to effectively address the right user groups and (2) identifying more vulnerable user cohorts. We also address existing challenges in identifying users’ vulnerabilities in understudied contexts like visual-based media and instant messaging services. In the end, as different users possess different vulnerabilities, we suggest that the ability to identify vulnerable users would allow interventions to effectively address the right group of users.

REFERENCES

[1] Mihai Avram, Nicholas Micallef, Sameer Patil, and Filippo Menczer. 2020. Exposure to social engagement metrics increases vulnerability to misinformation. Harvard Kennedy School Misinformation Review (Jul 2020). https://doi.org/10.37016/mr-2020-033

[2] Samantha Bradshaw and Philip N Howard. [n. d.]. Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. ([n. d.]), 26. https://demtech.oi.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf
[26] Myrto Pantazi, Scott Hale, and Olivier Klein. 2021. Social and Cognitive Aspects of the Vulnerability to Political Misinformation. Political Psychology (2021). https://doi.org/10.1111/pops.12797

[27] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A Machine Learning Approach to Twitter User Classification. In Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011, Lada A. Adamic, Ricardo A. Baeza-Yates, and Scott Counts (Eds.). The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2886

[28] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. 2020. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychological Science 31, 7 (2020), 770–780. https://doi.org/10.1177/0956797620939054 arXiv:https://doi.org/10.1177/0956797620939054 PMID: 32683243.

[29] Francisco Rangel, Anastasia Giachanou, Bilal Hisham Hasan Ghanem, and Paolo Rosso. 2020. Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on twitter. In CEUR Workshop Proceedings, Vol. 2696. Sun SITE Central Europe, 1–18. https://riunet.upv.es/handle/10251/166528

[30] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In Proceedings of the 32nd ACM Conference on Hypertext and Social Media (Virtual Event, USA) (HT ’21). Association for Computing Machinery, New York, NY, USA, 189–199. https://doi.org/10.1145/3465336.3475101

[31] Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. 2020. Susceptibility to misinformation about COVID-19 around the world. Royal Society open science 7, 10 (2020). https://doi.org/10.1098/rsos.201199

[32] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451807

[33] Kai Shu, Deepak Mahadeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv preprint arXiv:1809.01286 (2018).

[34] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding User Profiles on Social Media for Fake News Detection. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). 430–435. https://doi.org/10.1109/MIPR.2018.00092

[35] Herbert A Simon. 1957. A behavioral model of rational choice. Models of man, social and rational: Mathematical essays on rational human behavior in a social setting (1957), 241–260.

[36] Michael Sülflow, Svenja Schafer, and Stephan Winter. 2019. Selective attention in the news feed: An eye-tracking study on the perception and selection of political posts on Facebook. New Media & Society 21, 1 (Jan. 2019), 168–190. https://doi.org/10.1177/1461444818791520

[37] Calum Thornhill, Quentin Meeus, Jeroen Peperkamp, and Bettina Berendt. 2019. A Digital Nudge to Counter Confirmation Bias. Frontiers in Big Data 2 (2019). https://doi.org/10.3389/fdata.2019.00011

[38] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559

[39] Fabiana Zollo and Walter Quattrociocchi. 2018. Misinformation Spreading on Facebook. Springer International Publishing, Cham, 177–196. https://doi.org/10.1007/978-3-319-77332-2_10