International Journal of Data Science and Analytics (2022) 13:265–269
https://doi.org/10.1007/s41060-022-00325-0

EDITORIAL

# Online information disorder: fake news, bots and trolls

**Anastasia Giachanou[1] · Xiuzhen Zhang[2] · Alberto Barrón-Cedeño[3] · Olessia Koltsova[4] · Paolo Rosso[5]**

Published online: 9 May 2022


**Abstract**
Recent years have seen a tremendous increase in the propagation of different types of misinformation and disinformation, including among others fake news, rumours, clickbait and conspiracy theories. Misinformation involved in satire and clickbait, among others, has a different intention from disinformation. Despite many attempts by the research community, the development of technology to assist experts in detecting disinformation remains an open problem due to a number of challenges. For example, there are various biases such as confirmation bias and peer pressure that hinder users from recognizing non-credible information. In addition, fake news is intentionally written to confuse the readers, often containing a mixture of false and real information. To this end, in this editorial we present current challenges in the area of fake news identification and discuss contributions published in our special issue.

**Keywords** Information disorder · Fake news · Fake news spreaders

## 1 Introduction

Recent years have seen a tremendous increase in the public interest to the spreading of different types of misinformation and disinformation, including fake news, rumours, and conspiracy theories. While misinformation is usually understood as generally unintentionally erroneous or inaccurate content, disinformation is defined as a message having an intention to deceive or mislead [21]. Neither represents a new phenomenon, but several important global changes have recently contributed to a faster dissemination of both truthful and untruthful content, and to the visibility of this process that has, in turn, led to public concern. The major factor is the proliferation of social media: content production and spreading has been taken away from professional media, which is supposedly bound to journalistic standards, relevant legislation and deontological ethics. Instead, social media have directly connected users with each other stimulating user-to-user communication, and to the actors whose contact to the audience had been previously mediated by journalists: advertisers, politicians and both local and foreign governments. The fall of these barriers has been coupled with technical affordances, such as the automation of content dissemination and even production, and the global character of many social media platforms.

The spread of fake news, rumours and conspiracy theories leads to severe consequences in different spheres, including economy, politics, and health. The COVID-19 pandemic has seen the rise of infodemics characterized by the viral spread of both misinformation and disinformation. For instance, the circulation of misinformation and conspiracy theories about the coronavirus disease has led to distorted information regarding its origin, prevention, and effective treatment [12]. More specific, health-related misinformation and conspiracy

✉ Anastasia Giachanou
a.giachanou@uu.nl

Xiuzhen Zhang
xiuzhen.zhang@rmit.edu.au

Alberto Barrón-Cedeño
a.barron@unibo.it

Olessia Koltsova
olessia.koltsova@gmail.com

Paolo Rosso
prosso@dsic.upv.es

1 Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

2 School of Computing Technologies, RMIT University, Melbourne, Australia

3 Università di Bologna, Bologna, Italy

4 National Research University Higher School of Economics, Moscow, Russia

5 PRHLT, Universitat Politècnica de València, Valencia, Spain

theories have disrupted the vaccination rates and promoted dubious treatment practices [10].

Politics has been particularly prone to intentional disinformation campaigns whose purpose has been either sewing division to provoke internal or international tensions, or mobilization around a particular way of intended action which, ultimately, contributes to polarization as well, by dividing social layers of a single society or nations into settings of hostile confrontations. Among other things, fake news stories have been heavily recognized as having a significant impact on the reputation of political personas, parties and countries, on the results of various elections and referendums [1,2], and have sometimes contributed to the start or legitimization of military conflicts. Moreover, these processes have been increasingly trans-border: national elites who were protected from disinformation attacks within national media systems, they are now more vulnerable to such attacks from international peer rivals and their globally penetrating media resources. Elites' concerns have become a major driver of broader public concerns with fake news proliferation.

Despite increasing efforts of the research community, the development of disinformation research is still in its cradle. This domain involves multitude tasks. The first and most obvious is the automatic detection of false information and related phenomena, such as automated information spreaders (bots), accounts prone to fake dissemination and accounts prone to one-sided ("half-true") information dissemination. The level of success of this task is connected to two related research goals. The first one is studying misinformation dissemination, including its speed, channels, audiences, pervasiveness, and other features whose understanding may help to increase the efficiency of fake news detection and prevention. The second one is studying the way humans perceive misinformation, including the factors of susceptibility, engagement and the cues and approaches humans use to tell truths from lies. Special emphasis should be put on what could be considered fake news. This is a key aspect to preserve critical thinking against the dogmatic thinking imposed by politicians ruling a country, or aiming at ruling it. The independent work of fact-checkers is crucial to distinguish critical thinking from disinformation strategies.

This special issue aims at covering all these domains of research, with a special focus on the main goal of fake news detection. The goal of this editorial is to provide an overview of the open research challenges for the community of researchers interested in analysing and modelling fake news, as well as an introduction to the special issue. The rest of the article is structured as follows. Section 2 discusses research challenges relevant in the context of this special issue. Section 3 briefly introduces the articles included in this JDSA special issue. Section 4 concludes the editorial.

## 2 Challenges

The complex human–machine interaction form with online misinformation and disinformation brings significant challenges to their automated detection and mitigation at scale, especially on social networking platforms.

Progress in disinformation detection is hampered by topic diversity, multi-modality and multi-linguality. Topics subject to disinformation range from politics to health and from natural disasters to celebrity news. Fact-checking services focus mostly on verifying political claims and require costly human annotation. Automated detection systems can scale the detection of disinformation, but their generalizability across topics is limited by the origin of training data. Disinformation is present not only in natural language text but also other media such as images and video clips. Research on multimedia disinformation detection systems is in its infancy, possibly due to the limited amount of the multimedia resources necessary to train machine learning models [6,7]. Multilinguality represents another significant challenge to the automated detection of disinformation. For instance, Twitter supports posting in over 40 languages, whereas most automated detection systems are designed for one language: English. This trend neglects the fact that large amounts of fake news spread in multiple languages across the globe to diverse demographic communities and therefore are more destructive.

Machine-generated or neural fake news [23] brings new challenges. NLP technologies have been developed to the stage where texts in fluent natural language can be generated, which can be misused by malicious actors to generate disinformation at scale. Research has shown that humans trust such machine-generated news more than human-authored news, which is alarming [5,23]. Different from human-authored fake news, machine-generated ones have the potential of being produced in mass at a low cost, both temporal and economic. Detection systems for human-authored fake news might not be effective against machine-generated ones.

Despite the tremendous efforts by fact checking services and automated detection systems, fake news still spread widely, as evidenced by the infodemic during the ongoing COVID-19 pandemic. Therefore, mitigation is important to confine the spread of disinformation. The traffic volume in social networks together with their dynamic information spreading and sparse links present significant challenges for designing effective strategies. Network level mitigation aims to run mitigation campaigns to spread corrective information widely on social networks to contain disinformation, which can focus on optimizing intensity for posting corrective information [4] or selecting optimal debunkers to spread corrective information [22]. Personal-level mitigation strategies have also been proposed [20]. Moreover, it is also desirable to design mitigation strategies considering the presence of adversaries, which may also spread disinformation to

counteract debunkers. Limited resources of real-world, open social networks presents significant challenges to conduct large-scale live experiments that are necessary to evaluate mitigation strategies.

Humans play a crucial role in the widespreading of disinformation on social media platforms. Limited research has been reported towards understanding the human behaviour for such spread. Early research focuses on understanding factors affecting human perception towards the credibility of online information [11], and characteristics such as gender, education, background and personal experience all affect an individual's perception [16]. Shortage of data on human–information and human–human interaction presents a challenge to more deeply understand the human trust towards misinformation, to identify influencers for propagating misinformation and corrective information, as well as to understand that formation of communities for misinformation on different domains and events.

There are still many open challenges for online disinformation management in terms of transparency, privacy and ethics. Transparency is especially important for automated disinformation detection systems. Explaining classification decisions certainly enhances the transparency and credibility of systems. Neural models are known as blackbox systems. Although the attention mechanism is leveraged to open the blackbox, it is still far from the objective of human-understandable explanation. Protection of user privacy is another challenge. The automated profiling of users and detection of rumour spreaders may inadvertently reveal personal information. Moreover, detection of fake news spreaders without fine-grained analysis can also raise ethical concerns. Members of the general public may unintentionally spread unverified or even false information from questionable sources. These users are amenable to stigmatization if they are identified as spreaders of false information.

The grand challenges for online information disorder detection and mitigation call for cross-disciplinary research from computer science, social science and journalism disciplines to work together to deepen our understanding of the complexity of human–machine interaction and develop systems to empower citizens to proactively identify and debunk false information, so as to ultimately quench online false information.

## 3 In this special issue

In this section, we briefly introduce each of the articles that were selected for this special issue. These contributions are a varied representation of the challenges and solutions that are emerging in this research context. Interestingly, some of the contributions focused on health-related fake news either from the perspective of an overview paper or from user study perspective [15,19,24]. On the other side, conspiracy theories were analysed in a study by Batzdorfer et al. [3]. Other contributions aimed to detect fake news by proposing methods based on Transformers [14,18]. Additionally, the role of users in the context of fake news was explored by Shrestha and Spezzano [17] and Parmentier et al. [13].

Disasters, health, and politics are three domains that attract a high number of fake news. To this end, Sadiq and Mathew reviewed 28 articles relevant to those topics and discussed how fake news spreads in those sectors [15]. In their overview paper, they also discussed the methodologies that have been used by researchers, theoretical perspectives, and strategies to control the spread of fake news.

Vlasceanu and Coman conducted two user studies to assess which information source could increase knowledge about COVID-19 and COVID-19 vaccines [19]. In their study, the participants were asked to rate the accuracy of 26 statements. The statements were assigned randomly to one of the following 10 between-subjects conditions that covered varying sources providing belief-relevant information: a political leader (Trump/Biden), a health authority (Fauci/CDC), an anecdote (Democrat/Republican), a large group of prior participants (Democrats/Republicans/Generic), or no source (Control). For the next phase, the participants were asked to rate the accuracy of the statements again. The design of the second study was similar with the difference that focused on the COVID-19 vaccine. The results of the two user-based studies showed that knowledge increased most when the source of information was a generic group of people, regardless of participants' political affiliation. Also the second study showed that expert communications were most successful at increasing Democrats' vaccination intentions. On the other side, Republicans' vaccination intention was not increased with any of the sources.

Zuo et al [24] focused as well on health-related fake news. In their study, they built a dataset with markers of credibility in online news articles and which addressed the two following challenges: (i) identify what is worth checking, in the specific context of a marker, and (ii) whether the perception provided by the marker to the general reader is indeed true. During the data collection and annotation, they found that even when news articles have explicit citations, they may sometimes mislead about medical findings.

Conspiracy theories are a special type of disinformation and could be defined as the belief that hidden coalitions of powerful individuals follow an agenda that intends or causes harm to society, the particular in-group of the individual, or the individual specifically. Batzdorfer et al [3] focused on conspiracy theories and analysed Twitter data across 11 months in 2020 from the timelines of 109 conspiracy theory posters. Also they analysed a comparison group (non-conspiracy theory group) of equal size. For the analysis, they used word embeddings not only to distin-

guish conspiracy and non-conspiracy theory content but also to identify which element of conspiracies emerged in the pandemic. In addition, Batzdorfer et al [3] applied time series analyses on the aggregate and individual level to investigate whether there is a difference between conspiracy and non-conspiracy posters as well as the temporal dynamics of the conspiracy tweets. The results of the study showed that there was a substantially higher level of overall COVID-19-related tweets in the non-conspiracy group and higher level of random fluctuations. Also, there was a positive trend in the conspiracy tweets and an increase in users in 2020. The aggregate series of conspiracy content revealed two breaks in 2020 and a significant albeit weak positive trend since June. On the individual level, the series showed strong differences in temporal dynamics and a high degree of randomness and day-specific sensitivity.

Two of the main challenges in fake news detection are their early detection and the unavailability or shortage of labelled data for training purposes. To address those two challenges, Raza and Ding [14] proposed a fake news detection framework that exploits the information from the news articles and the social contexts to detect fake news. The proposed model is based on a Transformer architecture, consisting of two parts: the encoder part that can learn useful representations from the fake news data, and the decoder part that predicts the future behaviour based on past observations. They also incorporated many features from the news content and social contexts into the model. In addition, Raza and Ding [14] proposed an effective labelling technique to address the label shortage problem. The results of their experiments showed that the model can detect fake news with higher accuracy within a few minutes after it was propagated (early detection) than the baselines.

Transformers and other modern neural language models can also be used by malicious actors to automatically produce textual content looking as it has been written by humans. The study by Stiff and Johansson [18] evaluated transformer-based detection algorithms in a large variety of experiments involving both in-distribution and out-of-distribution test data, as well as evaluation on more realistic in-the-wild data. The results showed that the generalizability of the detectors can be questioned, especially when applied to short social media posts. Moreover, the experiments showed that the best performing detector (RoBERTa-based) was non-robust even to basic adversarial attacks, demonstrating how easy it is for malicious actors to remain undetected by the current state-of-the-art detection algorithms.

Users' role in disseminating fake news has become inevitable with the increase in popularity of social media as a means of news source. People in social media actively participate in the creation and propagation of news, favouring the proliferation of fake news intentionally or unintentionally. Thus, it is necessary to identify the users who tend to share fake news to mitigate the rampant dissemination of fake news over social media. Shrestha and Spezzano [17] focused on that problem by performing a comprehensive analysis on two Twitter datasets and investigating the patterns of user characteristics in social media in the presence of misinformation. Specifically, they studied the correlation between the user characteristics and their likelihood of being fake news spreaders and demonstrated the potential of the proposed features in identifying fake news spreaders. The proposed approach achieved an average precision ranging between 0.80 and 0.99 on the considered datasets, consistently outperforming baseline models. Furthermore, in line with previous studies [8,9] they also showed that the user personality traits, emotions, and writing style are strong predictors of fake news spreaders.

Parmentier et al [13] focused also on users but from a different perspective. In particular, they proposed a data-driven multi-faceted trust modelling which incorporated many distinct features and demonstrated how clustering of similar users enables a critical new functionality: supporting more personalized, and thus more accurate predictions for users. The proposed framework was evaluated with a trust-aware item recommendation task in the context of a large Yelp dataset. Moreover, they discussed how improving the detection of trusted relationships in social media can help online users to fight the spread of misinformation and rumours, within a popular social networking environment.

## 4 Conclusions

In this editorial, we discussed current challenges in the area of fake news and we presented the contributions of the special issue. The contributions covered various research topics in the area of disinformation, ranging from user behaviour to the detection of fake news with neural networks. We hope that the contributions of the special issue can be useful for further research in this field.

**Acknowledgements** We would like to thank the Editor-in-Chief, Longbing Cao, and all the authors for their valuable contributions. In addition, we are deeply grateful to the reviewers that helped with the decision process and contributed with excellent reviews to make this issue possible.The work of Paolo Rosso was in the framework of the Iberian Digital Media Research and Fact-Checking Hub (IBERIFIER) funded by the European Digital Media Observatory of the EC (2020-EU-IA-0252), and of the XAI-DisInfodemics research project on eXplainable AI for disinformation and conspiracy detection during infodemics (PLEC2021-007681), funded by MCIN/AEI and by European Union NextGenerationEU. The work of Anastasia Giachanou was funded by the Dutch Research Council (grant VI.Vidi.195.152).

## References


1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. J. econ. perspect. **31**(2), 211–36 (2017)
2. Bastos, M.T., Mercea, D.: The brexit botnet and user-generated hyperpartisan news. Soc. Sci. comp. Rev. **37**(1), 38–54 (2019)
3. Batzdorfer, V., Steinmetz, H., Biella, M., et al.: Conspiracy theories on twitter: Emerging motifs and temporal dynamics during the covid-19 pandemic. Int. J. Data Sci, Analyt (2022)
4. Farajtabar, M., Yang, J., Ye, X., et al.: Fake news mitigation via point process based intervention. In: International Conference on Machine Learning, PMLR, pp 1097–1106 (2017)
5. Gagiano, R., Kim, M.M.H., Zhang, X.J., et al.: Robustness analysis of grover for machine-generated news detection. In: Proceedings of the The 19th Annual Workshop of the Australasian Language Technology Association, pp 119–127 (2021)
6. Giachanou, A., Zhang, G., Rosso, P.: Multimodal fake news detection with textual, visual and semantic information. In: Proceedings of the 23rd international conference on text, speech and dialogue, pp 30–38 (2020)
7. Giachanou, A., Zhang, G., Rosso, P.: Multimodal multi-image fake news detection. In: Proceedings of the 7th IEEE international conference on data science and advanced analytics, pp 647–654 (2020)
8. Giachanou, A., Rosso, P., Crestani, F.: The impact of emotional signals on credibility assessment. J. Assoc. Inform. Sci. Technol. **72**(9), 1117–1132 (2021)
9. Giachanou, A., Ghanem, B., Rissola, E., et al.: The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. Data& Knowledge Eng. 138 (2022)
10. Loomba, S., de Figueiredo, A., Piatek, S.J., et al.: Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the UK and USA. Nat. Human Behav. **5**(3), 337–348 (2021)
11. Morris, M.R., Counts, S., Roseway, A., et al.: Tweeting is believing? understanding microblog credibility perceptions. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, pp 441–450 (2012)
12. Organization, W.H., et al.: Coronavirus disease (covid-19) advice for the public: Myth busters 30 (2020)
13. Parmentier, A., Cohen, R., Ma, X., et al.: Personalized multi-faceted trust modeling to determine trust links in social media and its potential for misinformation management. Int. J. Data Sci, Analytics (2022)
14. Raza, S., Ding, C.: Fake news detection based on news content and social contexts - a transformer based approach. Int. J. Data Sci, Analytics (2022)
15. Sadiq, T.M., Mathew, S.K.: The disaster of misinformation: a review of is research in social media. Int. J. Data Sci, Analytics (2022)
16. Shariff, S.M., Zhang, X., Sanderson, M.: On the credibility perception of news on twitter: readers, topics and features. Comp. Human Behav. **75**, 785–796 (2017)
17. Shrestha, A., Spezzano, F.: Characterizing and predicting fake news spreaders in social networks. Int. J. Data Sci, Analytics (2022)
18. Stiff, H., Johansson, F.: Detecting computer-generated disinformation. Int. J. Data Sci, Analytics (2022)
19. Vlasceanu, M., Coman, A.: The impact of information sources on covid-19 knowledge accumulation and vaccination intention. Int. J. Data Sci, Analytics (2022)
20. Wang, S., Xu, X., Zhang, X., et al.: Veracity-aware and event-driven personalised news recommendation for fake news mitigation. In: Proceedings of the 2022 ACM Web Conference (2022)
21. Wardle, C., Derakhshan, H.: Information disorder: Toward an interdisciplinary framework for research and policymaking (2017)
22. Xu, X., Deng, K., Zhang, X.: Identifying cost-effective debunkers for multi-stage fake news mitigation campaigns. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pp 1206–1214 (2022)
23. Zellers, R., Holtzman, A., Rashkin, H., et al.: Defending against neural fake news. Adv. Neural Inform. Process. Syst. 32 (2019)
24. Zuo, C., Mathur, K., Kela, D., et al.: Beyond belief: A cross-genre study on perception and validation of health information online. Int. J. Data Sci, Analytics (2022)