IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/47ISM/2026/4727826/
Vol. 2026 (7), Article ID 4727826

Selecting AI Solutions for Managing False Information During Disasters: A Prescriptive Recommendation Framework*

Jérémie KATEMBO KAVOTA, Daniel TOMIUK, Yasmeen ABUHASIRAH and Hamid SHIRAZIAN

Université du Québec à Montréal, Canada

Correspondence should be addressed to: Daniel TOMIUK, tomiuk.daniel@uqam.ca

*Presented at the 47th IBIMA International Conference, 29-30 June 2026, Madrid, Spain

Abstract

Social media has become an essential communication tool during disaster situations. Unfortunately, it is also used to spread false information very rapidly. This includes rumors, outdated or misleading information, doctored images, and fake news stories. These can undermine and disrupt public trust and emergency response efforts and further endanger already impacted populations. Although research has addressed how AI can identify false information, much of it has focused on individual AI models. There remains a lack of prescriptive guidance to assist disaster management teams in selecting appropriate AI configurations for various types of disasters and false information. To address this, we propose a framework using a design science research approach. This paper reviews literature on disaster-related false information, AI-enabled misinformation detection, multimodal analysis, real-time adaptive models, cloud-based AI, federated learning, and Edge AI. We then provide a framework for three essential design tasks: mapping false information to disaster types, determining the factors that influence AI use in these situations, and recommending appropriate AI solutions. Our framework shows that different disaster contexts require different AI solutions. For instance, rapidly spreading text-based misinformation requires real-time natural language processing and network analysis, visual misinformation requires multimodal analysis and authenticity verification, privacy-sensitive disasters may require federated learning or Edge AI, and recurring disasters may benefit from cloud-based AI that facilitates the use of shared datasets. Overall, this research proposes a prescriptive artifact to help disaster managers select appropriate AI solutions to manage misinformation on social media during disasters.

Keywords: Disaster management; social media; false information; cloud-based AI; federated learning; Edge AI

Introduction

Social media (SM from now on) has become a ubiquitous means of communication in the context of disaster management. Access to quick and relevant information regarding the disaster is essential for effective disaster management and responding to its impacts. Despite the numerous benefits that SM platforms provide during disasters, they can also be a source of false and inaccurate information. These can come in many forms. Rumors, “fake news”, and other types of unverified information shared online can cause confusion, severely undermining public safety and relief efforts (Hilberts et al., 2024; Muhammed & Mathew, 2022).

Because SM platforms allow for massive amounts of information to be quickly disseminated by numerous people during disaster situations, false information can be difficult to detect especially when it appears in short messages such as tweets. SM can also convey falsehoods indirectly through misleading wording, satire, exaggeration, or selective omissions. It is often difficult to distinguish this from legitimate information (Saxena et al., 2020).

Cite this Article as: Jérémie KATEMBO KAVOTA, Daniel TOMIUK, Yasmeen ABUHASIRAH and Hamid SHIRAZIAN Vol. 2026 (7) “ Selecting AI Solutions for Managing False Information During Disasters: A Prescriptive Recommendation Framework " Communications of International Proceedings, Vol. 2026 (7), Article ID 4727826, https://doi.org/10.5171/2026.4727826

Artificial intelligence (AI) techniques, including methods like machine learning (supervised and unsupervised), deep learning, and natural language processing (NLP) have been created as a means of flagging false information shared on SM (Aïmeur et al., 2023; Bodaghi et al., 2024; Rajdev & Lee, 2015). This paper proposes a prescriptive recommendation framework to support decision-makers in selecting AI-based solutions for managing false information during disasters. Such a framework is important for several reasons.

First, false information tends to be very viral. Because it evokes strong negative emotions, false information usually spreads significantly faster than factual content (Vosoughi et al., 2018). Being able to detect and mitigate falsehoods is of utmost importance. Second, the algorithms that push content on SM platforms tend to prioritize messages that are engaging (i.e., that generate many "likes" or shares) regardless of message accuracy, further exacerbating the problem. Third, once false information is "out there", circulating and gaining traction, correcting it can become very challenging, so focus should be on early detection. Fourth, many of today's AI-powered detection systems are primarily text-based and often trained on narrow/locally specific crisis datasets. However, false information on SM can also spread using manipulated images, deepfake videos, and through audio recordings. Given these many potential sources for false information, appropriate AI-based solutions must be selected and often used in combination. Finally, crises are fast evolving. Detection models that fail to capture real-time data can also result in an incomplete representation of conditions on the ground.

The remainder of this paper is structured as follows. We begin with a literature review on what is meant by false information. We identify contextual, technical, and organizational factors that can influence the framework's recommendations. After that, we outline a research methodology using design science. Finally, the last section of the paper discusses the expected contributions of our framework and we conclude the paper.

While several AI techniques have been proposed for misinformation detection in the literature, most studies have focused on individual detection models, and there is a lack of a prescriptive framework to guide decision-makers in choosing appropriate AI solutions based on the characteristics of a disaster and its context, infrastructure constraints, and operational requirements. In this study, we address this gap by proposing a recommendation framework to help decision-makers choose AI configurations tailored to disaster scenarios.

This work contributes to the literature in three ways. First, it synthesizes disaster misinformation patterns and AI detection techniques into a decision-oriented framework. Second, it extends AI-enabled disaster management research by shifting attention from individual detection models to the selection of context-appropriate AI configurations. Third, it provides a design science artifact that can guide disaster managers in choosing among NLP, multimodal detection, network analysis, cloud-based AI, edge AI, federated learning, and hybrid solutions (McMahan et al., 2017; Kairouz et al., 2021).

Literature review

Social Media and the Spread of False Information

Misinformation is defined as false information that is spread without any malicious intent but instead because of personal beliefs and emotions (Sandu et al., 2024; Wang et al., 2025). Individuals who share misinformation believe it to be true (Wardle & Derakhshan, 2017). Misinformation often arises through confusion, fear and information that is incomplete and out-of-date. For instance, misinformation shared on Social Media (SM) platforms can include an out-of-date weather forecast report that is sharing incorrect information about the movement of storms. Other types of misinformation include spreading incorrect information about evacuation procedures and where to find the resources needed to survive the storm. Information of this type can lead victims to travel to areas that are both further into the storm and where the resources are not actually available (Sandu et al., 2024; Wang et al., 2025).

Disinformation, on the other hand, is false information shared intentionally. The objective is to deceive or harm (Wardle & Derakhshan, 2017; Bodaghi et al., 2024). Finally, *fake news* is a specific subset of disinformation. It refers to fabricated stories pretending to be legitimate or coming from legitimate sources. The intent of fake news is to mislead recipients of the message for political, financial, or social agendas (Tandoc, Lim, & Ling, 2018). Individuals with malicious intentions have the potential to perform actions that make it difficult for those attempting to perform rescues to do their jobs, to discredit the efforts of the government (especially if those in power are of rival political parties), or to spread lies regarding their efforts or management of the situation altogether. Fraudsters may also attempt to take advantage of the emotions and feelings of urgency that individuals will have regarding the disaster and its effects upon those experiencing it, encouraging individuals to donate to funds and charities that are fake.

Beyond the intentional falsehoods shared on SM platforms during disasters are the unintentional falsehoods that threaten lives of disaster victims. The spread of misinformation and disinformation on social media puts individuals in danger by diminishing trust in the authorities and the response to the disaster, and puts their lives in danger (Hilberts et al., 2024).

The ubiquity of SM platforms and their algorithms prioritizing engagement are likely to have a compounding effect. False information can evoke strong negative emotions such as anxiety, stress and fear, emotions which likely drive individuals to share information more impulsively, amplifying information diffusion across large audiences. With SM platforms, false information can "go viral" very quickly before any type of corrective message or action can be sent or taken.

The sheer volume of both accurate and inaccurate information on SM can further undermine disaster response efforts because of conflicting information. Moreover, being bombarded with a continuous flow of updates (some true, others not) is likely to contribute to information overload. This makes it much more difficult to discern reliable and relevant information from falsehoods, leaving both victims and authorities struggling to process critical details needed to make life-saving decisions (Houston et al., 2015).

Leveraging AI for Detecting False Information

AI helps identify and mitigate false information on SM (Wang et al., 2025). AI detection falls into three categories: *content-based*, *social and structured-based*, and *hybrid* techniques (Bondielli & Marcelloni, 2019; Chen et al., 2021). Content-based detection focuses on the content generated by users. Social and structure-based detection analyzes patterns of dissemination of false information, and the role users play in spreading and amplifying it through network dynamics. Hybrid detection combines content-, user-, and network-based features. In these three strategies, *machine learning* is predominant. Learning techniques include *supervised* learning, where models are trained using labeled data to recognize patterns and *unsupervised* learning, which finds hidden patterns in data without predefined labels. *Deep* learning, a more advanced form of machine learning, uses layers of artificial neurons to analyze complex information such as images, videos and long textual data. These techniques can work together to detect false information by spotting misleading content and tracking how it spreads (Chen et al., 2021). Advanced Natural Language Processing (NLP) models, such as BERT, RoBERTa (Robustly Optimized BERT), and GPT (Generative Pre-trained Transformers) can recognize anomalies and deceptive language that aims to cause panic (e.g., "The authorities are lying to us") or by identifying unsubstantiated sources using certain linguistic markers (e.g., "some people say", "reportedly", "allegedly", "apparently", "I've heard that"). The use of these types of markers within the content allows for NLP models to recognize instances of potentially false information without requiring any additional contextual information.

AI can also help validate information across multiple languages and SM platforms (Bodaghi et al., 2024). Verification techniques include *misinformation tracking* to understand how it evolves, spreads through networks (tracking its engagement and share rate), and affects public opinion and behavior. *Stance classification* analyzes the attitude expressed in text messages by classifying it as supportive, neutral, or opposing. It analyzes sentiment and positioning of the sender and can predict the author's viewpoint on a given topic. Additionally, verification can employ *multitask classification* techniques to accurately determine whether a post contains disputed factual claims, which remains quite challenging for natural language processing algorithms (Metaxas et al., 2015). Because false information cannot be entirely eliminated, response efforts increasingly focus on reducing its impact through detection, verification, and timely intervention.

Different kinds of disasters tend to give rise to particular types of false information. During natural disasters (e.g., hurricanes, floods), false information shared on SM often takes the form of outdated information, exaggerated claims regarding the source and severity of the disaster, false claims regarding resource availability (e.g., incorrect location details, misinformation on the availability of aid or shelter), conspiracy theories, and scams perpetrated by predatory individuals (Houston et al., 2015). During pandemics, it is common to receive false medical advice, unfounded theories regarding its origins, and discreditation of health officials (Shu et al., 2017). Whereas in man-made disasters, such as terrorist attacks, false information aims to achieve political or ideological objectives. The intention is to spread fear throughout a population and to create distrust towards its government (Shu et al., 2017).

AI Capable of Processing Video, Audio and Multimedia Content

Many AI-powered false information detection models analyze textual content (i.e., after the disaster), which limits their scalability and accuracy and can introduce algorithmic bias. Another limitation of focusing only on the text

of the posts is that visual and multimedia content is often highly influential in the spread of false information. Studies show that humans tend to trust visual content over textual content, considering it to be more factual. Therefore, doctored images and videos are often more likely to spread on social media platforms, and more difficult to refute (Zhou & Zafarani, 2020).

Beyond text-based detection, there is a critical need for multimodal machine learning (Liang et al., 2024). Researchers are developing these systems to combat the issue of misinformation across different content formats. Deep learning models like BERT, GPT, ViTs for images, and MesoNet for videos can analyze large amounts of data very quickly and allow AI to understand how different types of data relate to one another. This strengthens misinformation detection by cross-checking different media types.

The Nature of the Disaster and Types of False Information

Research shows that different types of disasters tend to be associated with particular patterns of false information circulating on SM. Authors like Houston et al (2015) and Sandu et al. (2024) state that different crisis situations tend to produce particular false information. For example, for natural disasters (e.g., hurricanes, earthquakes), the false information published is often of the type of information that becomes outdated quickly due to the fast-changing nature of the crisis. The misinformation usually concerns location and disaster severity. People share incorrect storm paths, previous evacuation routes, and incorrect updates about the availability of resources (food, water and medical aid). During outbreaks and pandemics (such as Covid-19 and Ebola), false information usually includes false medical advice (unverified treatments and prevention methods), the spread of conspiracy theories regarding the origins of the outbreak, and false information targeting health officials and the government (Shu et al., 2017).

In man-made disasters (e.g., terrorist attacks), false information can take the form of fake news such as implicating particular groups of people or exaggerating casualty numbers or the claim of additional attacks to spread further fear and panic throughout the population. Other man-made disasters such as power outages, industrial accidents or even technological disasters tend to spark unsubstantiated speculations as to the causes of such a disaster, falsely suggesting it was caused by the incompetence of the government or even some malicious activity without actual evidence of such an occurrence. The discovery of such disasters often elicit fake instructions from individuals as to the best way to respond to the disaster, some of which may suggest actions that are unsafe to take in the face of the crisis (Shu et al., 2017).

Given the role that the nature of the disaster plays in the type of false information that is published regarding those disasters, a prescriptive recommendation framework could assist in determining which AI techniques can be utilized to mitigate the false information in those specific types of disasters.

Real-time, Adaptive AI Models

AI-based filtering of false information on SM has traditionally relied on retrospective text-based content analysis. This approach has drawbacks. Analyzing content much after it has already spread, allows false information to circulate and gain momentum, spreading uncontrollably before corrective actions can be taken, potentially hampering disaster response (Shu et al., 2017).

To address this, researchers have developed real-time, adaptive AI models. These continuously monitor information as it is being shared. They use network analysis to track how false information spreads across SM, identify key influencers and can prioritize interventions. Effectiveness is improved through continuous refinements based on real-world communications, processed in real-time rather than retrospectively (Shu et al., 2017). Integrating multimedia analysis (e.g., image, video, audio) also helps detect more sophisticated forms of false information, such as deepfakes and other manipulated content, which are likely to be missed with only text-based analyses (Zhou & Zafarani, 2020). This significantly improves the speed and accuracy of misinformation detection and the response adequacy during disasters.

Another key feature of AI-based misinformation detection systems is explainability, and decision-makers are reluctant to make decisions based on AI recommendations if they do not understand why the system makes the recommendation; explainable AI (XAI) techniques can increase transparency so that emergency managers can better understand the rationale for why some content is identified as misinformation and why some AI

configurations are recommended to increase trust, accountability, and adoption of AI in disaster management contexts (Adadi & Berrada, 2018).

Although adaptive and real-time AI models improve response times, this increased performance comes at a cost. They require considerable computing resources because continuously monitoring and processing large volumes of data requires advanced IT infrastructure and increased storage capacities. These requirements significantly raise costs (due to their complexity, as well as their development, testing and implementation challenges). This affects their applicability and practicality in regions with limited technological infrastructure, such as in developing countries. Even when IT infrastructure and budgets are available, adaptive and real-time AI solutions might be excessive. Not all disasters will require such sophisticated AI-based solutions, particularly when the crisis is localized, SM activity is limited, or existing response infrastructure is sufficient. Smaller, more localized disasters with fewer SM users may not warrant the costs and efforts required to put into place real-time adaptive AI systems to curb the dissemination of false information.

Local vs. Shared Datasets and Models

Training AI using locally gathered disaster data can be slow because these are limited to patterns observed in a specific context. Shared datasets expose models to a broader range of disaster scenarios. This improves their ability to detect emerging misinformation patterns. This is possible when AIs contribute to a cloud computing database of information about disasters from around the world. Each of these AIs will gain knowledge at a faster rate through the sharing of information, leading to the AIs gaining the ability to improve their skills in handling various disaster situations.

By leveraging cloud-based AI solutions, disaster response systems can become more scalable, adaptable, and effective in its aims of learning from the data collected from various disasters worldwide and enhancing the disaster response strategy more rapidly. Cloud-based AI systems can collect, store, and analyze data from disasters worldwide and feed that information to AI systems in the local area. Furthermore, machine learning models can learn from the data collected from disasters worldwide to enhance the accuracy of the disaster response for local areas even before there are similar disasters of that nature in those specific locations. AIs can better respond regionally and to new kinds of disasters through globally shared data and learning models. This technology can rapidly collect disaster data coupled in real-time via IoT devices in disaster-prone areas and share it with other AI.

Cloud-based AI enables centralized data-sharing but raises data security and privacy concerns. *Federated Learning* addresses this by training AI models on decentralized data without sharing raw data, ensuring better privacy while enabling collaborative learning. *Edge AI* and related distributed architectures can reduce latency and bandwidth dependence by processing data closer to where it is generated (Aboualola et al., 2023). Analyzing data locally in real time avoids cloud over-dependency by transmitting only essential insights and not raw high-volume data. Federated Learning and Edge AI are therefore privacy-preserving alternatives to the cloud, optimizing AI performance through decentralized processing (McMahan et al., 2017; Kairouz et al., 2021).

Definition of the Design Artifact

The design artifact proposed in this paper is a **prescriptive recommendation framework** intended to support disaster managers in selecting appropriate AI-based solutions for managing false information on SM during disasters. The artifact is not itself a misinformation detection algorithm. Rather, it is a framework that will assist in the selection of appropriate AI techniques and architectures for the various types of disaster situations.

The artifact takes as its inputs a variety of parameters related to the type of disaster occurring, the type of false information that is being spread, the modality of the content that is being disseminated, the urgency of responding to that false information, and the availability of relevant datasets and resources. Based on these inputs, the artifact generates recommendations regarding suitable AI configurations, such as text-based NLP detection, multimodal detection, network-based rumor tracking, real-time adaptive AI, cloud-based AI, edge AI, federated learning, or hybrid configurations.

The purpose of this artifact is therefore to translate these various conditions into a decision about which AI solutions will be best suited to address the described disaster. This can aid decision-makers in avoiding generic and poorly matched AI solutions to disaster scenarios.

Design Science Framework for the Recommendation System

Building on the literature reviewed above, this paper proposes a prescriptive recommendation framework to support the selection of AI solutions for managing false information during disasters. The framework is organized around three design tasks. First, it maps disaster types to likely false-information patterns. Second, it identifies contextual, technical, and organizational factors that influence the suitability of different AI techniques and deployment architectures. Third, it translates these factors into a recommendation logic that can guide decision-makers toward appropriate AI configurations.

Table 1 presents the input–process–output logic of the proposed recommendation framework. The framework takes disaster-specific, informational, technical, and organizational factors as inputs, applies a structured decision logic, and generates recommendations regarding suitable AI techniques and deployment architectures.

Table 1. Input-Process-Output Logic of the Proposed Recommendation Framework

Component	Description
Inputs	Disaster type, false-information type, content modality, speed of diffusion, data availability, infrastructure capacity, privacy sensitivity, need for real-time response
Decision logic	Rule-based or expert-validated mapping between contextual conditions and suitable AI capabilities
Outputs	Recommended AI techniques and architectures, such as NLP, multimodal detection, network analysis, real-time adaptive AI, cloud AI, edge AI, federated learning, or hybrid solutions
Users	Disaster managers, emergency-response organizations, public-sector crisis communication teams
Expected value	More context-sensitive, timely, and operationally feasible AI selection for managing false information

Table 2 illustrates how the framework translates disaster-specific and infrastructural conditions into AI configuration recommendations.

Table 2. Illustrative AI configuration recommendations

Contextual condition	Recommended AI configuration
Text-heavy misinformation with rapid diffusion	Real-time NLP + network analysis
Visual/video manipulation	Multimodal detection + human verification
High privacy sensitivity	Federated learning or edge AI
Limited infrastructure	Lightweight/local model or edge AI
Cross-regional recurring disaster	Cloud-based AI + shared datasets

The following example illustrates how the framework can be used in practice. In the case of a hurricane, some of the false information that could be spread may include information regarding incorrect evacuation routes, incorrect information regarding the locations of evacuation shelters, the dissemination of manipulated media of areas that are flooding in the hurricane, or the spread of information regarding road closures. Each of these inputs can be utilized by the framework to determine appropriate countermeasures to those types of false information. It could then recommend real-time NLP monitoring, geolocation-based verification, multimodal image analysis, and a cloud-edge architecture to support rapid local processing while drawing on broader disaster-related datasets. This example illustrates how the framework translates disaster-specific conditions into AI configuration recommendations.

Following the design science research process, this study focuses primarily on problem identification, objective definition, artifact design, and preliminary demonstration. Full empirical evaluation is beyond the scope of this paper, but the framework is designed to be evaluated in future work through expert validation, scenario-based simulations, and retrospective application to past disaster cases.

To develop the prescriptive recommendation framework, we employ a design science approach organized around three stages. Our framework covers both the theoretical foundations behind the artifact, including the mapping of false-information patterns and the identification of influencing factors, and the practical implementation and validation of the recommendation system. Here are the proposed stages:

(1) Map the types of false information for the different types of disasters: This essential stage forms the basis of the framework's recommendation logic. By understanding the patterns of false information that emerge during specific types of disasters, it will also be possible to understand which AI systems are best for detecting and countering those specific types of false information. Mapping these patterns will make the recommendation logic more accurate. This stage will need to include a comprehensive review of literature on disaster-specific false information. Data from past disasters and SM analysis can be used to refine the mapping.

(2) Identify other factors that impact the choice of an AI system for reducing false information: As highlighted in previous sections, AI selection does not only depend on the type of false information. This stage aims to ensure that the recommendation framework takes a broader view of AI suitability. Factors such as regional characteristics (e.g., SM penetration), disaster characteristics (e.g., scale of the disaster), cost of the AI solution, and the speed at which false information disseminates should also be considered when deciding on the type of AI and whether it will be real-time adaptive or a simpler retrospective model. A thorough literature review and interviews with AI experts will help identify these factors. Aside from those mentioned previously, they will likely include technical factors (e.g., computational resources required by the AI, whether there is a need to process multimedia content, and whether real-time data feeds are available). Organizational factors are also likely to influence the framework's recommendations. This can include the level of expertise available and existing infrastructure in place when implementing the AI and how robust it is.

(3) Develop the Recommendation Framework Using a Design Science Approach: Design science (DS) primarily focuses on developing artifacts to address real world problems (Tuunanen et al., 2024). The process includes identifying the problems, creating a solution (or artifact), testing the solution's effectiveness (usually by experts), and often creating new iterations of the artifact after receiving feedback. The DS framework proposed by Peffers et al. (2007) includes six specific steps for creating innovative solutions to the problems identified in the practical world. These steps are (1) *Problem Identification and Motivation*, where a real-world issue is identified; (2) *Define Objectives for a Solution,* where goals for the solution are established; (3) *Design and Development,* which involves creating the actual artifact aimed at solving the problem; (4) *Demonstration* where the artifact is showcased to decision-makers; (5) *Evaluation* which requires a thorough assessment of the artifact's performance; and (6) *Communication*, where results and insights are shared with various stakeholders (e.g., academia, disaster management personnel). This process ensures that the solution is iteratively improved based on feedback and evaluation of experts in step 5.

The framework's usefulness can be evaluated through pilot testing with emergency managers and AI practitioners. Simulations of various disaster events can also be conducted. Alternatively, data from past disasters can be used to examine whether the framework recommends AI configurations that are appropriate for specific disaster scenarios. Future research should evaluate the framework through expert review, scenario-based simulations, and retrospective application to past disaster events.

Conclusion

Social media is an important communication channel used in disaster management. However, there are also challenges to using social media. One is the spread of false information during disaster events on social media channels. This paper proposes a prescriptive recommendation framework to help decision-makers select suitable AI solutions for managing false information on social media during disaster situations. The framework takes into account the various aspects of the complexity of disasters, such as the type of disaster, the false information that is spread, the need for multimodal detection, and the requirements for real-time responses to disasters, as well as the optimal use of the available technologies. By implementing it as a recommendation system, the framework could help decision-makers in customizing AI solutions according to the specific characteristics of disaster-management contexts.

Acknowledgment

The authors used generative AI tools, specifically ChatGPT 5.5, to support editing, language refinement, and the identification of grammatical and semantic issues. AI tools were not used to generate or alter the paper's core ideas or scholarly contribution. The authors reviewed and approved all AI-assisted revisions and remain fully responsible for the manuscript's content.

References

Aboualola, M., Abualsaud, K., Khattab, T., Zorba, N. and Hassanein, H.S. (2023) 'Edge technologies for disaster management: A survey of social media and artificial intelligence integration', IEEE Access, 11, pp. 73782–73802. https://doi.org/10.1109/ACCESS.2023.3293035

Adadi, A. and Berrada, M. (2018) 'Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)', *IEEE Access*, 6, pp. 52138–52160. https://ieeexplore.ieee.org/document/8466590

Aïmeur, E., Amri, S. and Brassard, G. (2023) 'Fake news, disinformation and misinformation in social media: A review', *Social Network Analysis and Mining*, 13(1), article 30. https://doi.org/10.1007/s13278-023-01028-5

Bodaghi, A., Schmitt, K.A., Watine, P. and Fung, B.C.M. (2024) 'A literature review on detecting, verifying, and mitigating online misinformation', *IEEE Transactions on Computational Social Systems*, 11(4), pp. 5119–5145. https://doi.org/10.1109/TCSS.2023.3289031

Bondielli, A. and Marcelloni, F. (2019) 'A survey on fake news and rumour detection techniques', *Information Sciences*, 497, pp. 38–55. https://doi.org/10.1016/j.ins.2019.05.035

Chen, X., Zhou, F., Zhang, F. and Bonsangue, M. (2021) 'Catch me if you can: A participant-level rumor detection framework via fine-grained user representation learning', *Information Processing & Management*, 58(5), article 102678. https://doi.org/10.1016/j.ipm.2021.102678

Hevner, A. and Chatterjee, S. (2010) *Design research in information systems: Theory and practice*. New York: Springer.

Hilberts, S., Evers, S., Govers, M. and Petelos, E. (2024) 'The impact of misinformation on social media in the context of natural disasters', *European Journal of Public Health*, 34(Supplement 3), ckae144.245. https://doi.org/10.1093/eurpub/ckae144.245

Houston, J.B., Hawthorne, J., Perreault, M.F., Park, E.H., Goldstein Hode, M., Halliwell, M.R. and Griffith, S.A. (2015) 'Social media and disasters: A functional framework for social media use in disaster planning, response, and research', *Disasters*, 39(1), pp. 1–22. https://doi.org/10.1111/disa.12092

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R. et al. (2021) 'Advances and open problems in federated learning', Foundations and Trends in Machine Learning, 14(1–2), pp. 1–210. https://doi.org/10.1561/2200000083

Liang, P.P., Zadeh, A. and Morency, L.-P. (2024) 'Foundations & trends in multimodal machine learning: Principles, challenges, and open questions', *ACM Computing Surveys*, 56(10), pp. 1–42. https://doi.org/10.1145/3656580

McMahan, B., Moore, E., Ramage, D., Hampson, S. and Aguera y Arcas, B. (2017) 'Communication-efficient learning of deep networks from decentralized data', in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR*, 54, pp. 1273–1282.

Metaxas, P.T., Finn, S. and Mustafaraj, E. (2015) 'Using TwitterTrails.com to investigate rumor propagation', in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. New York: ACM, pp. 69–72. https://doi.org/10.1145/2685553.2702691

Muhammed T, S. and Mathew, S.K. (2022) 'The disaster of misinformation: A review of research in social media', *International Journal of Data Science and Analytics*, 13(4), pp. 271–285. https://doi.org/10.1007/s41060-022-00311-6

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007) 'A design science research methodology for information systems research', *Journal of Management Information Systems*, 24(3), pp. 45–77.

Rajdev, M. and Lee, K. (2015) 'Fake and spam messages: Detecting misinformation during natural disasters on social media', in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE, pp. 17–20. https://doi.org/10.1109/WI-IAT.2015.102

Sandu, A., Ioanăș, I., Delcea, C., Geantă, L.-M. and Cotfas, L.-A. (2024) 'Mapping the landscape of misinformation detection: A bibliometric approach', *Information*, 15(1), article 60. https://doi.org/10.3390/info15010060

Saxena, A., Hsu, W., Lee, M.L., Leong Chieu, H., Ng, L. and Teow, L.N. (2020) 'Mitigating misinformation in online social network with top-k debunkers and evolving user opinions', in *Companion Proceedings of the Web Conference 2020*. New York: ACM, pp. 363–370. https://doi.org/10.1145/3366424.3383297

Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017) 'Fake news detection on social media: A data mining perspective', *ACM SIGKDD Explorations Newsletter*, 19(1), pp. 22–36. https://doi.org/10.1145/3137597.3137600

Tandoc, E.C., Lim, Z.W. and Ling, R. (2018) 'Defining "fake news": A typology of scholarly definitions', *Digital Journalism*, 6(2), pp. 137–153. https://doi.org/10.1080/21670811.2017.1360143

Tuunanen, T., Winter, R. and vom Brocke, J. (2024) 'Dealing with complexity in design science research: A methodology using design echelons', *MIS Quarterly*, 48, pp. 427–458.

Vosoughi, S., Roy, D. and Aral, S. (2018) 'The spread of true and false news online', *Science*, 359(6380), pp. 1146–1151. https://doi.org/10.1126/science.aap9559

Wang, J., Zhai, Y. and Shahzad, F. (2025) 'Mapping the terrain of social media misinformation: A scientometric exploration of global research', *Acta Psychologica*, 252, article 104691. https://doi.org/10.1016/j.actpsy.2025.104691

Wardle, C. and Derakhshan, H. (2017) *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Strasbourg: Council of Europe. Available at: https://rm.coe.int/information-disorder-report-november-2017/1680764666

Zhou, X. and Zafarani, R. (2020) 'A survey of fake news: Fundamental theories, detection methods, and opportunities', *ACM Computing Surveys*, 53(5), pp. 1–40. https://doi.org/10.1145/3395046