Social Media Data: Challenges, Mitigation Strategies, and Opportunities for Disaster Management

Samia Abid
Data Science, AI & Modelling Centre (DAIM)
University of Hull
Hull, United Kingdom
samia.abid-2022a@hull.ac.uk

Bhupesh Kumar Mishra
DAIM
University of Hull
Hull, United Kingdom
bhupesh.mishra@hull.ac.uk

Usama Younis
Department of CS&IT
MUST, AJK, Pakistan
muhammadusamabhatti009@gmail.com

Dhavalkumar Thakker
School of Computer Science
University of Hull
Hull, United Kingdom
D.Thakker@hull.ac.uk

Nishikant Mishra
Hull University Business School
University of Hull
Hull, United Kingdom
Nishikant.Mishra@hull.ac.uk

***Abstract*—For disaster management, the accurate and timely availability of factual information is crucial for effective decision-making and response. Traditional communication channels are either costly or do not provide real-time data. Here comes the role of social media data due to its ability to swiftly disseminate real-time information, allowing for community engagement and situational awareness of the disasters. However, leveraging social media data for this application is not without challenges. It is quite challenging to deal with the four Vs-volume, velocity, variety, and veracity of the social-media data to increase its value for decision-making. To overcome these challenges and enhance the quality of data, implementing robust techniques and strategies is of profound importance. This paper highlights critical challenges associated with using social media data in disaster management and proposes a comprehensive framework of methods, techniques, algorithms, and methodologies to improve data quality. By systematically addressing these challenges, we can harness the full potential of social media data to support disaster response efforts and ultimately save lives.**

***Index Terms*—Social-media Data, Disaster Management, Disaster Response, Data Quality, Data Reliability.**

I. Introduction

The unexpected eruption of emergencies requires timely and accurate disaster management to control and evade its consequences. During times of disasters, people move towards emergency management organizations and helplines like 999 and 911 [1]. Due to a high number of rescue requests, it is not practical for these authorities to respond to the requests of all the victims. In such situations, social media platforms have been proven to be valuable sources of real-time information [2].

Social media platforms like Facebook, Twitter, Instagram, YouTube, etc, bridge the communication gap between people and help to spread information across the globe. With technological advancements and internet accessibility, people are more into social media. They share their opinions, information, and sentiments in the form of text, images, audio, and videos with their acquaintances and the public [3].

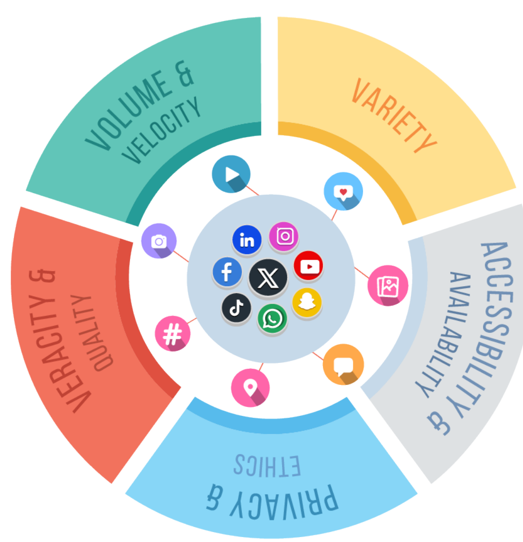

Fig. 1. Unstructured Social Media Data and its Challenges

Due to the availability of personal gadgets a myriad of users, including victims and volunteers, approach social media during natural or man-induced calamities to connect to their families, post rescue requests, provide situation awareness updates, and share information about the needs of victims [3]. Whereas traditional data sources like cameras [4], RFID readers [5], and GPS information [6] have been utilized in the literature for disaster management. Social media data has several promising merits over the data mentioned in earlier sources. It provides the real-time availability of user-generated content that cannot be acquired through traditional data sources and government and regulatory organizations. Additionally, scrapping data is quicker and more economical than traditional data sources. Moreover, conventional data sources are prone to malfunction during natural disasters like earthquakes, floods, cyclones, etc; social media remains unaffected [3].

Yet, there is a need to address several challenges of unstructured and uncertain social media data and improve data quality to be employed for sensitive decision-making relevant to lives at risk during disasters (Figure 1. It is challenging to deal with the four Vs-volume, velocity, variety, and veracity of the social-media data to increase its value for decision-making. To overcome these challenges and enhance the quality of data, it is of profound importance to combat issues, i.e., misinformation, multimodal and multi-lingual data handling, information overload, sarcasm detection, biases identification, misinterpretations of data, privacy and ethical concerns, accessibility and availability of social media data. This article is dedicated to identifying potential challenges of social media data and highlighting the approaches, tools, algorithms, and strategies to overcome these challenges and improve data quality.

The rest of the paper is organized as follows: Section II highlights the key challenges of social media data. Section III provides the strategies and solutions to improve data quality. Section IV concludes the work.

II. Challenges of Social Media Data for Disaster Management

This section details the challenges of the unstructured and uncertain nature of social media data. A pictorial representation of the challenges is shown in Figure 1. While Figure 2 provides a taxonomy of areas requiring attention to improve social media data.

A. Data Veracity and Quality

The integrity and quality of social media-based disaster-relevant information pose several challenges to decision-making and response. Trustworthiness is a serious concern for disaster management organizations. Incorrect information (i.e., misinformation, fake news, rumors, etc.) results in improper utilization of resources for response efforts and situational awareness. The literature emphasizes certain features to assess the authenticity of disaster-relevant posts [7]. These features comprise the utilization of URLs, the count and credibility of followers, and grammar correctness. Despite these features, upholding trust in social media information remains largely unresolved. For instance, posts/tweets may contain intentionally misleading information covertly done, sometimes obviously, to gain more views and comments. For example, during the COVID-19 disease outbreak, untrue information was shared regarding locations, people, treatment measures, and the number of casualties. In this case, the candidate's

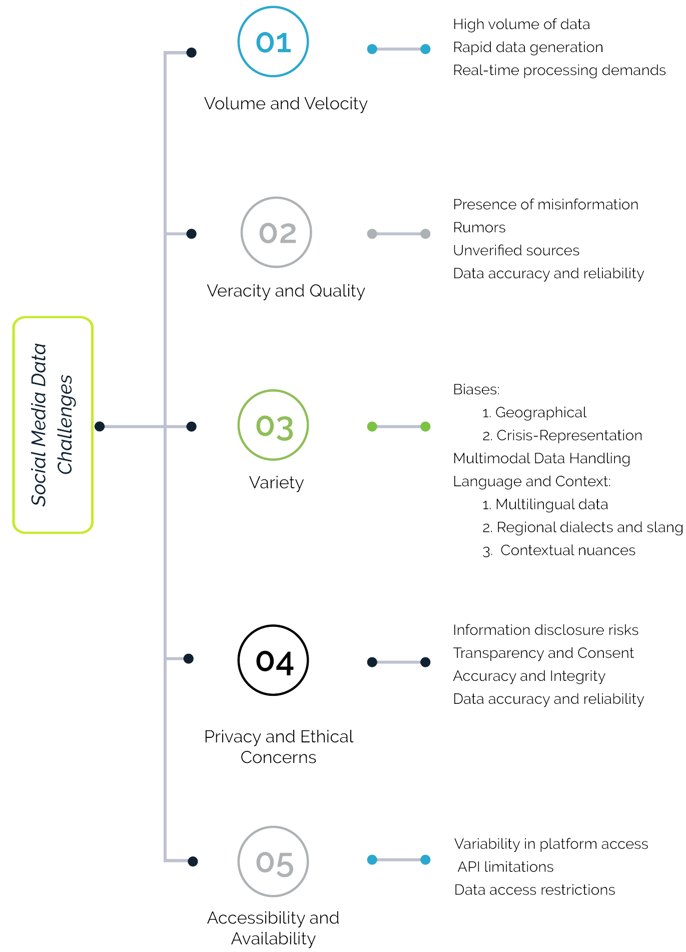

Fig. 2. Taxonomy of areas requiring attention to improve social-media data

proposed use of the number of comments and likes to test the validity of posts/tweets may not be effective [8]. Moreover, a tweet may be sarcastic in which it may not mean what it portrays but the opposite [9]. It is also a possibility that all tweets are not from humans. Bots (automated) accounts are presented on the X platform which may have been purpose-tailored for disaster-related hashtags and generating tweets that may not correctly report a real natural disaster event or situation at the peak of such event naturally occurring and getting traction [10].

The spread of misinformation due to insufficient information is another aspect that affects the integrity of information It is driven by the anxiety suffered by disaster-affected communities [11]. Individuals turn to social media platforms due to the lack of authentic information from government authorities on their official pages. This conduct can reduce attention but contributes to the spread of unverified information.

B. Data Variety

During a disaster, government officials and the public share various informative situational updates on social media. Due to the human-centric nature of social media data, it may be prone to issues relevant to different types of biases resulting from the skewed data distribution concerning the geographical and crisis-representation distribution. Moreover, user-generated content is shared in many languages and formats (i.e., text, videos, images, etc.). For effective and responsible disaster management, it is mandatory to eliminate these biases resulting from the real-time uncertainties of social media data. Otherwise, utilizing social media data without dealing with these concerns may result in unfair responses and management. The challenges occurring due to its inherent variety are as follows:

- *Crisis, Location, and People Representation Biases:* It is noticed that social media users do not pay much attention to crises in less populated or less-known regions. Social media is biased towards urban users and regions. Potential bias in the data-gathering process may be possible due to inequalities in access to the internet. The underrepresentation of rural areas and older participants skews the data towards a particular group of people [11] [12] , locations [13], or disaster types. Hence, designing a social media-based generalized disaster management system becomes complicated.
- *Multimodal Data Handling* Data on social media platforms like Twitter, Facebook, YouTube, Instagram, etc., is generated in many forms, including text, images, videos, and geological data. [7]. Each data format requires different processing approaches; hence, the diversity of data makes it cumbersome to analyze and integrate it for decision-making [14]. The integration and fusion of various types of data is a complex process that requires advanced tools and techniques.
- *Language and Context* Social media users from different geographical locations post helpful information in their native language other than English [15]. This information cannot be overlooked because it may provide important information and require a swift response. Hence, the linguistic diversity in social media increases the complexity of natural language processing. The user-generated content may be in different dialects and languages or contain slang, abbreviations, sarcasm, or sentiments. In such scenarios, analyzing the multilingualism element and the content context becomes problematic during disasters [7].

Although social media provides valuable insights for disaster management, its variety must be carefully management to ensure accurate analysis.

C. Data Volume and Velocity

The early detection of natural disasters using social media data is of paramount importance [16]. However, the vast amount of data and velocity of the data in disastrous situations make it difficult to identify the trends, events, and topics in unstructured social media communications [17]. The high volume of data necessitates scalable approaches for data collection and analysis which require real-time processing to extract insightful information from the user-generated content [18]. This issue further complicates the extraction of meaningful information from the unstructured content which is prone to grammatical errors, sarcasm, and multilingual information. Furthermore, data visualization is an essential step for emergency management organizations to save lives by quickly acting [17]. In the disaster mitigation phase, victims and volunteers share information about the disasters. With the increase in data volume conventional tools no longer handle a vast amount of data to be visualized that too in different formats and languages [19].

D. Data Accessibility and Availability

One of the basic principles of scientific research is to reproduce the existing findings of the researchers by utilizing the same datasets and methods from published articles. However, many big datasets are not publicly available which limits the reproducibility of such articles. Due to the legal constraints of the Twitter API, the Twitter dataset cannot be forwarded to anyone other than their research group. Moreover, the high cost of Twitter data is also a major challenge to restrict access to social media data. The data access problem is a challenge that can impede the utilization of social media data in the future [13]. For instance, datasets like CrisisNLP, TweetDIS, Disaster Response Data (DRD), and CrisisLex datasets require contacting the author or may be available through research collaborations. To combat this issue many researchers have developed social media crawlers/ scrappers to gain access to the social media datasets. However, due to the privacy updates to the social media platforms, the scrappers are obsolete because social media platforms implement measures to prevent automated scrapping. Thus, the restrictions on sharing data also reduce the replicability of data analysis [13].

E. Data Privacy and Ethical Concerns

To ensure the user privacy of social media data it is required to obtain user consent and adhere to ethical implications, especially concerning the location data [20]. The study in [21] highlights the ethical implications of employing social media data and expresses his concern regarding the data protection of social media users. X platform requires that tweets be republished only in their original form and attributed to the original poster. Since this can lead to the re-identification of the poster, paraphrasing and anonymization may not be the options here [22]. Hence, this study [23] emphasizes the need for coherent consent action plans to address privacy and ethical concerns.

III. Strategies and Solutions

Several useful tools and libraries have been introduced to address the challenge of social media data. This section provides the techniques to combat the challenges highlighted in the previous section.

First, we provide tools and techniques to deal with the volume and velocity of social media data for disaster detection and response applications. Various platforms like Apache

TABLE I
TOOLS TO HANDLE LARGE VOLUMES AND VELOCITY OF SOCIAL MEDIA DATA

Challenges	**Tools**	**Description**
Data Analysis	BigSheets BigInsights [25]	Helps to analyze data from social networks.
Analysis and Visualization	Apache Hadoop Distributed [25]	Designed to address the huge data analysis, computation, and visualization problem.
Collection and Aggregation	Apache Flume [25]	Developed by Cloudera primarily for online analytics. Assists in data collection and aggregation.
Data Analysis and Visualization	Tableau and Rapid Miner [28]	Tools for analyzing, evaluating, and visualizing Snapshot's data.
Data Visualization	Power BI [27]	For visual analytics of social media data.
Real-Time Data Processing	Amazon Kinesis [26]	Assists in collecting, processing, and analyzing real-time streaming data.
Big Data Analytics	MapReduce and Apache Spark [24]	Fast and easy-to-use processing of large-scale data.

Spark [24] and Apache Hadoop [25] can play a vital role in processing a huge volume of social media data; while Apache Spark can offer high performance for real-time data analytics. Furthermore, to handle a high volume of data, Amazon Kinesis services can enable and retrieve real-time data from social media [26]. Moreover, Microsoft Power BI and Tableau can be used to establish interactive dashboards [27]. RapidMiner is another tool that assists in the clustering and predictive analysis of huge volumes of data [28]. Additionally, graph visualization libraries are useful for understanding social media data. These libraries include D3.js, Gephi, etc [27]. These tools and libraries can help to overcome the issue of large volumes of social media data and help to monitor, process, analyze, and timely respond to disaster-relevant data. Further details regarding these tools are provided in Table I.

Further to ensure the reliability of social media posts for decision-making in disaster-relevant scenarios, it is mandatory to first authenticate the information and assess its quality. Social media content may be used to spread misleading/false information. To prevent the spread of such fake news and approve the credibility of social media content literature highlights numerous machine learning and transformer-based techniques as provided in Table II. Moreover, various other tools and techniques have been developed as provided in Table III. These tools help to assess the authenticity of social media data and improve its quality for further processing.

Removing biases from the dataset is crucial to ensure a fair and effective disaster management and response system. This can be achieved through diverse sampling, including various regions, age groups, rural and urban areas, disaster types, and user-generated content in different languages. Techniques like SMOTE can help address the issue of skewed representativeness in the dataset. Moreover, handling multimedia data like text, images, etc needs attention. When working with text data, Natural Language Processing (NLP) techniques such as lemmatization, POS tagging, and named entity recognition (NER) play a crucial role in feature extraction. Pre-trained language models (LLMs) like GPT-3 or BERT can be beneficial in disaster management by extracting meaningful features from disaster-related text. Similarly, pre-trained Convolutional Neural Networks (CNNs) are useful for extracting features from disaster-related images [35]. Additionally, sentiment analysis and integrating text, image, and sentiment features can provide a better understanding of disaster situations, contributing to effective disaster response strategies. The presence of various languages in social media data can introduce prejudices that restrict the inclusivity and comprehensiveness of the content. To tackle language prejudices, multilingual data gathering and translation services can be utilized. Multilingual language models have been effective in addressing these prejudices. Notable instances of multilingual language models include mBERT [39], BLOOM [40], mT5 [41], Falcon [42], PaLM [43], and LLaMA [44]. However, challenges such as the quality of language data and inherent biases in language data persist. Achieving genuine multilingualism requires the creation of high-quality multilingual datasets, an area that necessitates more attention from researchers.

Further to the issue of accessibility and availability of social media data, social media platforms often enforce limitations on access to stop automated scraping and safeguard user privacy. These limitations include rate limits, and API access controls. Diffbot and ParseHub are tools that utilize advanced AI and visual extraction methods to bypass these restrictions and extract structured data. Scrapy and Beautiful Soup rely on traditional web scraping techniques to gather content from both static and dynamic pages, but they may struggle with more sophisticated anti-scraping measures. Selenium, used for automating browser interactions, is effective in scraping dynamic content that needs user interactions. WebHarvy simplifies the scraping process with its point-and-click interface, but it may encounter limitations due to platform security features. Each tool has its strengths in overcoming access restrictions, depending on the complexity of the target website and the type of data needed.

To ensure the privacy and ethics of using social media, users should comply with the ethical and privacy standards of social media sites. It is mandatory to review the licensing requirements of social media sites and comply with their terms of service. Users must avoid storing the raw data and ensure that data is analyzed consistently with social media site guidelines. Moreover, sample tweets should not be published so that people cannot identify individual users, and users' privacy be protected. Instead, using common phrases and data aggregation should be promoted to anonymize and protect the privacy of social media users.

IV. CONCLUSION

Disaster management requires swift and effective decision-making during times of crisis. Due to the limitation of traditional data sources, the world is moving towards social media

TABLE II
TOOLS AND TECHNIQUES FOR MISINFORMATION AND RUMOR DETECTION

Challenge	**Tools/Algorithms**	**Description**
Misinformation	CrowdTangle [30]	A tool by Meta to identify misinformation spread on social media.
Fake News Detection	AltNews, PolitiFact, BSDetector, APF Fact Check, Reverse Image Search, Snopes [31]	These tools assist in reviewing the credibility of content online.
Rumor Detection	Support Vector Machines (SVM) [32], Naïve Bayes [33], Random Forest [33], K-Nearest Neighbors (KNN) [32], Logistic Regression [33]	These machine-learning algorithms assist in rumor detection from social media data.
Bot Detection	Support Vector Machines (SVM), Random Forest, Neural Networks [34]	Machine learning-based bot detection of tweets' text content.
Rumor and Misinformation Detection	Convolutional Neural Networks (CNN) [35], Recurrent Neural Networks (RNN) [36], Graph Neural Networks (GNN) [37], Transformer Models [36], Hybrid Models [38]	Deep learning techniques for detecting misinformation and preventing rumor spread on social media.

TABLE III
TOOLS TO ENHANCE DATA QUALITY AND TRUST

Tool/Technique	**Reference**	Description
Hootsuite, Sprout Social	https://www.hootsuite.com/ https://sproutsocial.com/	Data aggregation and collection from social media platforms that help in accurate data analysis
OpenRefine, Alteryx	https://openrefine.org/ https://www.alteryx.com/	Clean and pre-process social media data to enhance reliability and usability.
Lexalytics, MonkeyLearn	https://www.lexalytics.com/ https://help.monkeylearn.com/	Sentiment analysis, public opinion mining, anomaly, misinformation, and biased content detection of social media posts.
TinEye, Google Reverse Image Search	https://tineye.com/ https://images.google.com/	Verify the authenticity of social media images and ensure the credibility of visual content
Snopes, FactCheck.org	https://www.snopes.com/ https://www.factcheck.org/	Rumour detection of social media posts and misinformation prevention.
Google Street View	Location Verification	Help to confirm the accuracy of location-based content.

platforms to seek help and post rescue requests during natural disasters like floods, hurricanes, earthquakes, etc. Social media platforms like Facebook, Twitter, Instagram, YouTube, etc, bridge the communication gap between people and help to spread information across the globe. However, with benefits, there are certain challenges of social media data. This research work provides a detailed analysis of social media data challenges. Overcoming these challenges is mandatory for disaster management and response research because it involves making decisions to save people's lives, securing the infrastructure of cities, and optimal resource allocation. Hence, dealing with challenges like data volume and velocity, handling a variety of data, improving data veracity and quality, abiding by ethical and privacy concerns, and dealing with access restrictions is mandatory in this area of practice. Hence, we provide a review of tools, strategies, and algorithms to combat these issues and help to ethically improve the quality of social media data.

Future improvements in identifying disasters using social media will need to focus on dealing with uncertain data and building trust by enhancing algorithms for filtering out irrelevant information and confirming accuracy. Combining live data sources with advanced analysis of public sentiments and mapping their geographical locations can increase precision. Furthermore, creating mechanisms for verifying information from the public and using data ethically will contribute to establishing reliability and trustworthiness in disaster-related information.

REFERENCES

[1] R. W. Aslam et al., "STRategies to manage Emergency ambulance Telephone Callers with sustained High needs: an Evaluation using linked Data (STRETCHED) - a study protocol," BMJ Open, vol. 12, no. 3, p. e053123, Mar. 2022, doi: 10.1136/bmjopen-2021-053123.

[2] L. Du, Y. Feng, L. Y. Tang, W. Kang, and W. Lu, "Networks in disaster emergency management: a systematic review," Aug. 01, 2020, Springer. doi: 10.1007/s11069-020-04009-5.

[3] A. Saroj and S. Pal, "Use of social media in crisis management: A survey," Sep. 01, 2020, Elsevier Ltd. doi: 10.1016/j.ijdrr.2020.101584.

[4] Asosiasi Pendidikan Tinggi Ilmu Komputer Indonesia, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, 2016 International Conference on Informatics and Computing (ICIC).

[5] T. Prakash, B. Pavithra, P. Devaprakash, B. Sivesh, and S. N. Gnanavel, "Enhancing Disaster Management with RFID Technology: Efficient Victim Data Tracing in Emergency Situations," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology, ICSEIET 2023, pp. 231–235, 2023, doi: 10.1109/ICSEIET58677.2023.10303547.

[6] Rasit. Eskicioglu, Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems. ACM, 2012.

[7] M. Imran, F. Ofli, D. Caragea, and A. Torralba, "Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions," Sep. 01, 2020, Elsevier Ltd. doi: 10.1016/j.ipm.2020.102261.

[8] E. Ferrara, S. Cresci, and L. Luceri, "Misinformation, manipulation, and abuse on social media in the era of COVID-19," J Comput Soc Sci, vol. 3, no. 2, pp. 271–277, Nov. 2020, doi: 10.1007/S42001-020-00094-5/FIGURES/1.

[9] J. G. C. M. Rader and B. Gamback, "Sarcasm Annotation and Detection in Tweets," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10762 LNCS, pp. 58–70, 2018, doi: 10.1007/978-3-319-77116-8_5.

[10] K. Hayawi, S. Saha, M. M. Masud, S. S. Mathew, and M. Kaosar, “Social media bot detection with deep learning methods: a systematic review,” Neural Comput Appl, vol. 35, no. 12, pp. 8903–8918, Apr. 2023, doi: 10.1007/S00521-023-08352-Z/FIGURES/4.

[11] S. Muhammed T and S. K. Mathew, “The disaster of misinformation: a review of research in social media,” May 01, 2022, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s41060-022-00311-6.

[12] M. H. Tsou, “Research challenges and opportunities in mapping social media and Big Data,” Cartogr Geogr Inf Sci, vol. 42, pp. 70–74, Aug. 2015, doi: 10.1080/15230406.2015.1059251.

[13] B. E. Lopez, N. R. Magliocca, and A. T. Crooks, “Challenges and opportunities of social media data for socio-environmental systems research,” Jul. 01, 2019, MDPI AG. doi: 10.3390/land8070107.

[14] D. Chakraborty, D. Rudrapal, and B. Bhattacharya, “A multimodal sentiment analysis approach for tweets by comprehending co-relations between information modalities,” Multimed Tools Appl, vol. 83, no. 17, pp. 50061–50085, May 2024, doi: 10.1007/S11042-023-17569-Y/FIGURES/8.

[15] S. Bazzaz Abkenar, M. Haghi Kashani, E. Mahdipour, and S. M. Jameii, “Big data analytics meets social media: A systematic review of techniques, open issues, and future directions,” Telematics and Informatics, vol. 57, Mar. 2021, doi: 10.1016/j.tele.2020.101517.

[16] J. Phengsuwan et al., “Use of social media data in disaster management: A survey,” Future Internet, vol. 13, no. 2, pp. 1–24, Feb. 2021, doi: 10.3390/fi13020046.

[17] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, “Social media analytics – Challenges in topic discovery, data collection, and data preparation,” Int J Inf Manage, vol. 39, pp. 156–168, Apr. 2018, doi: 10.1016/J.IJINFOMGT.2017.12.002.

[18] M. K. preprint arXiv:2307.04046 and undefined 2023, “Social media analytics in disaster response: A comprehensive review,” arxiv.orgM KarimiziaraniarXiv preprint arXiv:2307.04046, 2023•arxiv.org, Accessed: Aug. 16, 2024. [Online]. Available: https://arxiv.org/abs/2307.04046

[19] S. Saleem and M. Mehrotra, ”Emergent Use of Artificial Intelligence and Social Media for Disaster Management,” Lecture Notes in Networks and Systems, vol. 287, pp. 195–210, 2022, doi: 10.1007/978-981-16-5348-3_15.

[20] A. Ghermandi et al., “Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use,” One Earth, vol. 6, no. 3, pp. 236–250, Mar. 2023, doi: 10.1016/J.ONEEAR.2023.02.008.

[21] Dr. B. Dhiman, “Ethical Issues and Challenges in Social Media: A Current Scenario,” SSRN Electronic Journal, Mar. 2023, doi: 10.2139/SSRN.4406610.

[22] J. Tomás, D. Rasteiro, and J. Bernardino, “Data Anonymization: An Experimental Evaluation Using Open-Source Tools,” Future Internet 2022, Vol. 14, Page 167, vol. 14, no. 6, p. 167, May 2022, doi: 10.3390/FI14060167.

[23] M. S. Al-Zaman, A. Khemka, A. Zhang, and G. Rockwell, “The Defining Characteristics of Ethics Papers on Social Media Research: A Systematic Review of the Literature,” J Acad Ethics, vol. 22, no. 1, pp. 163–189, Mar. 2024, doi: 10.1007/S10805-023-09491-7/TABLES/12.

[24] S. Ibtisum et al., “A comparative analysis of big data processing paradigms: Mapreduce vs. apache spark,” https://wjarr.com/sites/default/files/WJARR-2023-2174.pdf, vol. 20, no. 1, pp. 1089–1098, Oct. 2023, doi: 10.30574/WJARR.2023.20.1.2174.

[25] M. Birjali, A. Beni-Hssane, and M. Erritali, “Analyzing Social Media through Big Data using InfoSphere BigInsights and Apache Flume,” Procedia Comput Sci, vol. 113, pp. 280–285, Jan. 2017, doi: 10.1016/J.PROCS.2017.08.299.

[26] A. Toshniwal, K. S. Rathore, A. Dubey, P. Dhasal, and R. Maheshwari, “Media Streaming in Cloud with Special Reference to Amazon Web Services: A Comprehensive Review,” Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, pp. 368–372, May 2020, doi: 10.1109/ICICCS48265.2020.9121097.

[27] Y. Wu, N. Cao, D. Gotz, Y. P. Tan, and D. A. Keim, “A Survey on Visual Analytics of Social Media Data,” IEEE Trans Multimedia, vol. 18, no. 11, pp. 2135–2148, Nov. 2016, doi: 10.1109/TMM.2016.2614220.

[28] N. Alsulaim, N. Alkhalaf, M. Alghamdi, R. Alolaqi, and H. Almughaidi, “Social Media Analytics using Visualization: Snapchat,” International Journal of Computer Networks and Communications Security, vol. 7, no. 5, pp. 91–95, 2019, Accessed: Aug. 17, 2024. [Online]. Available: www.ijcncs.org

[29] P. Rajapaksha, R. Farahbakhsh, and N. Crespi, “BERT, XLNet or RoBERTa: The Best Transfer Learning Model to Detect Clickbaits,” IEEE Access, vol. 9, pp. 154704–154716, 2021, doi: 10.1109/ACCESS.2021.3128742.

[30] S. Mohammad, A. Al Jobair, F. Shaiara, and S. Sharmin, “Bridging the Gap: A Case Study of Utilizing Social Media to Accelerate Recovery and Structuring Disaster Management,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 14705 LNCS, pp. 154–172, 2024, doi: 10.1007/978-3-03161312-8_11.

[31] S. Hangloo and B. Arora, “Fake News Detection Tools and Methods – A Review,” Nov. 2021, Accessed: Aug. 18, 2024. [Online]. Available: https://arxiv.org/abs/2112.11185v1

[32] A. Kumar and S. R. Sangwan, “Rumor Detection Using Machine Learning Techniques on Social Media,” Lecture Notes in Networks and Systems, vol. 56, pp. 213–221, 2019, doi: 10.1007/978-981-13-2354-6_23.

[33] V. U. Gongane, M. V. Munot, and A. Anuse, “Machine Learning Approaches for Rumor Detection on Social Media Platforms: A Comprehensive Survey,” Lecture Notes in Electrical Engineering, vol. 858, pp. 649–663, 2022, doi: 10.1007/978-981-19-0840-8_50.

[34] M. Heidari, J. H. J. Jones, and O. Uzuner, “An empirical study of machine learning algorithms for social media bot detection,” 2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021 - Proceedings, Apr. 2021, doi: 10.1109/IEMTRONICS52119.2021.9422605.

[35] A. Alsaeedi and M. Al-Sarem, “Detecting Rumors on Social Media Based on a CNN Deep Learning Technique,” Arab J Sci Eng, vol. 45, no. 12, pp. 10813–10844, Dec. 2020, doi: 10.1007/S13369-020-04839-2/TABLES/32.

[36] L. Tan, G. Wang, F. Jia, and X. Lian, “Research status of deep learning methods for rumor detection,” Multimed Tools Appl, vol. 82, no. 2, pp. 2941–2982, Jan. 2023, doi: 10.1007/S11042-022-12800-8/TABLES/8.

[37] B. Pattanaik, S. Mandal, and R. M. Tripathy, “A survey on rumor detection and prevention in social media using deep learning,” Knowl Inf Syst, vol. 65, no. 10, pp. 3839–3880, Oct. 2023, doi: 10.1007/S10115-023-01902-W/METRICS.

[38] M. Sadat Aothoi, S. Ahsan, N. N. Choudhury, and A. A. Rasel, “Supervised Hybrid Model for Rumor Classification: A Comparative Study of Machine and Deep Learning Approaches,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 14148 LNCS, pp. 281–286, 2023, doi: 10.1007/978-3-031-39831-5_25.

[39] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” [Online]. Available: https://github.com/tensorflow/tensor2tensor

[40] B. Workshop et al., “BLOOM: A 176B-Parameter Open-Access Multilingual Language Model,” Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.05100

[41] L. Xue et al., “mT5: A massively multilingual pre-trained text-to-text transformer,” Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.11934

[42] G. Penedo et al., “The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only,” Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.01116

[43] A. Chowdhery et al., “PaLM: Scaling Language Modeling with Pathways,” 2023. [Online]. Available: http://jmlr.org/papers/v24/22-1144.html.

[44] H. Touvron et al., “Llama 2: Open Foundation and Fine-Tuned Chat Models,” Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.09288