Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters.

Rosa Vicari[1] and Nadejda Komendatova[1]

[1]Cooperation and Transformative Governance Research Group, Advancing Systems Analysis Program at IIASA (International Institute for Applied Systems Analysis)

OBJECTIVE

Within the expanding and diverse area of studies dedicated to AI tools to deal with misinformation, scientists cannot discard the following question:

- **what kind of gatekeepers do we wish moderation and recommendation algorithms—and also social media users—to be?**

This question addresses fundamental human rights and journalism ethics, such as :

- freedom of expression,
- the right of the public to be informed,
- accuracy,
- differentiation between fact and opinion,
- privacy protection,
- hate and discrimination prevention,
- plagiarism prevention...

DATA & METHODS

To address this question, we carried out a **meta-analysis of studies published in Scopus and Web of Science**.

1. We extracted **668 papers** that contained keyterms related to the topic of "AI tools to deal with misinformation on social media during hazards and disasters."
2. First, we selected **13 review papers** to identify relevant variables and refine our research scope
3. Then we screened the rest of the papers and **identified 266 publications as being significant** for our research goals.

Corpus selection based on keyterms included in abstracts (Web of Science & Scopus)

Abstract =(disaster) OR (emergenc) OR (hazard) OR (disaster) OR (flood) OR (earthquake) OR (industrial accident) OR (terrorist attack*) OR (COVID) OR (pandemic) OR (wildfire) OR (Coronavirus)*

AND Abstract =(social media) OR (Twitter) OR (WhatsApp) OR (Facebook) OR (Instagram) OR (YouTube)

AND Abstract =(detect) OR (monitor) OR (prevent) OR (screen) OR (AI) OR (artificial intelligence)

AND Abstract =(fake news) OR (misinformation)

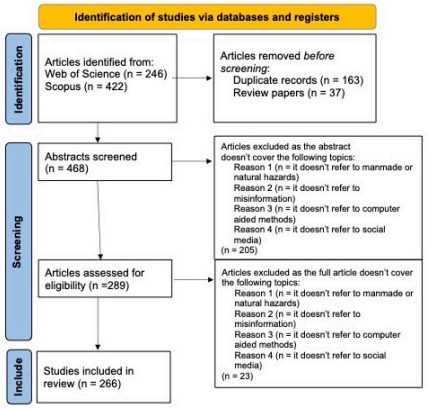

Figure 1. Our data selection process, guided by the PRISMA 2020 flow diagram, which outlines the steps involved in conducting a meta-analysis and the corresponding information flow. This diagram serves as useful tool for documenting the number of documents that were selected, assessed, deemed eligible or ineligible, as well as the reasons for exclusion (Page et al., 2021).

For each eligible paper, we analyzed **its objective, sponsor's location, year of publication, research area, type of hazard, and related topics.** As methods of analysis, we applied: **descriptive statistics, network representation of keyword co-occurrences, and flow representation of research rationale.**

REFERENCES

Bogart, S. (2022). SankeyMATIC.

Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. Telecommunications Policy, 39(9), 751–760. https://doi.org/10.1016/j.telpol.2014.12.003

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, n71. https://doi.org/10.1136/bmj.n71

Sarrica, M., Farinosi, M., Comunello, F., Brondi, S., Parisi, L., & Fortunati, L. (2018). Shaken and stirred: Social representations, social media, and community empowerment in emergency contexts. Semiotica, 2018(222), 321–346. https://doi.org/10.1515/sem-2016-0208

Varma, R., Verma, Y., Vijayvargiya, P., & Churi, P. P. (2021). A systematic survey on deep learning and machine learning approaches of fake news detection in the pre- and post-COVID-19 pandemic. International Journal of Intelligent Computing and Cybernetics, 14(4), 617–646. https://doi.org/10.1108/IJICC-04-2021-0069

INTRODUCTION

- **Social media** are supplanting traditional media as the **leading information source**.
- Social media contribute to the **social representation of disasters** (Sarrica et al., 2016):
  - They shape the population's perceptions and attitudes regarding disasters.
  - Hence **misinformation** can strongly affect disaster risk management.
- The traditional role of the journalist as **gatekeeper** is being undermined.
  - They don't hold anymore exclusive rights to the dissemination of news.
- Within the context of social media, **content recommendation algorithms** and **individual media users** serve as the new gatekeepers (Philip and Napoli, 2015)
- Machine Learning (ML) and Deep Learning (DL) are two popular approaches to automate the process of classifying news as fake or real (Varma et al., 2021).
- ML and DL algorithms are two subsets of the category of Artificial Intelligence (AI).

RESULTS 1

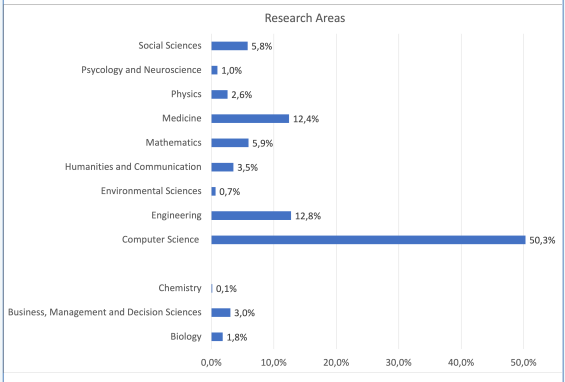

Figure 2. Chart showing the distribution of research areas in the corpus of studies. It is clear from the chart that the largest share of papers (50,3%) pertains to studies in the field of "Computer Science".

RESULTS 2

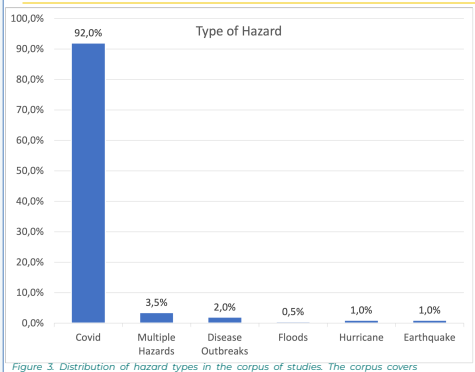

Figure 3. Distribution of hazard types in the corpus of studies. The corpus covers various types of hazards, and the figure displays the percentage of papers pertaining to each hazard. Notably, the vast majority of studies focus on COVID-19 (92%).

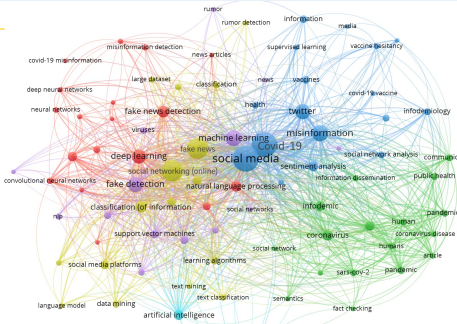

Figure 4. Visualization of the abstract dataset as a network. The author keywords and the index keywords are depicted as nodes, and their co-occurrence in each publication is shown as links between two nodes. One of the most prominent nodes in the network corresponds to the keyword "COVID-19". Each network cluster groups together keywords that frequently appear together. The red and blue clusters, with 21 and 15 nodes respectively, consist mainly of keywords related to the research scope. In contrast, the yellow and green clusters, with 15 and 20 nodes respectively, contain mostly keywords related to analytical methods, especially AI techniques.

RESULTS 3

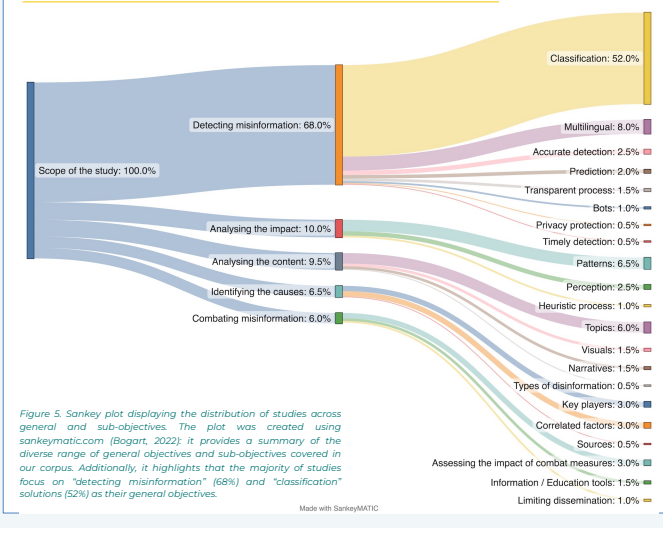

Figure 5. Sankey plot displaying the distribution of studies across general and sub-objectives. The plot was created using sankeymatic.com (Bogart, 2022): it provides a summary of the diverse range of general objectives and sub-objectives covered in our corpus. Additionally, it highlights that the majority of studies focus on "detecting misinformation" (68%) and "classification" solutions (52%) as their general objectives.

CONCLUSIONS & PERSPECTIVES

- Only **11%** of all publications are **social science** papers and **5%** are **decision science** papers.
  - These two research fields seem underrepresented for a topic is strongly connected to human reasoning.
- A minor portion of papers is dedicated to other than **COVID19 risks**.
- The majority of the studies is dealing with the question of **detecting misinformation**.
  - Is the decision to filter the news left to the convenience of individual users?
  - Are the individual users considered as active actors in attempts to combat misinformation?
  - Do researchers and practitioners have the same vision?

CONTACT INFORMATION

ROSA VICARI
vicari@iiasa.ac.at