Disinformation and Health: A systematic review study on health automatic fact-checking during the Pandemic

Haline Maia
INESC TEC and Faculty of Engineering of the University of Porto
Rua Dr. Roberto Frias, s/n
Porto, Portugal 4200-465
haline.c.maia@inesctec.pt

Extended Abstract

This research aims to improve the reflection on artificial intelligence and its potential to automate fact-checking activities, from operational and relational aspects to the final services, adopting a technological approach. Despite the importance of technical solutions, as Arnold (2020) points out, credibility problems are associated with fact-checking systems within only automatic phases. Therefore, a viable approach to address this challenge is to employ a human-in-the-loop solution, allowing human fact-checkers to leverage automated systems (Nakov et al. 2021). Surprisingly, there is little research conducted in this specific direction, and this research rightly aims to help encourage themes that may serve this task. Within this, the study seeks as its central point to raise issues that have been little explored in the automation of verified health-related information, which could be examined in more detail in future multidisciplinary research.

The paper will present an overview of emerging and related research areas in fact-checking, focusing on a systematic literature review of computational approaches. It also delves into the broad fact-checking pipeline and its essential elements, such as check-worthiness estimation, detecting fact-checked claims, stance detection, source reliability estimation, and identifying malicious social media content during the pandemic. While automation is deemed helpful for fact-checking, evidence is unclear about its use in disseminating results to the final users, particularly in debunking health-related hoaxes. As a new proposal for an old question, the recommendation is to adopt a more robust interdisciplinary collaboration between journalists and artificial intelligence developers. The challenges faced by traditional communication companies need to be addressed considering multiple angles. This work aims to provide a collaborative perception of automation initiatives and their applications in specialized topical domains.

Fact-checking is a process that involves verifying the accuracy of information. However, this process is not always straightforward, as research has shown that even among professional fact-checkers, there is disagreement one out of every five times (Miller 2020). The study by Vlachos and Riedel (2014) describes the following typical sequence of fact-checking steps: (a) extracting statements to be fact-checked, (b) constructing questions, (c) obtaining evidence from relevant sources, and (d) reaching a verdict based on that evidence. Commonly, this process takes many hours or days, and by that time, misleading statements may have spread out of control.

Studies from Full Fact (Babakar and Moy 2016) also present the automatic phases in four subtasks. The leading subtask is (1) monitoring, which involves extracting possible claims. Later, in subtask (2) spotting, the most numerous claims are selected. The (3) checking subtask analyzes the share corresponding to its fairness, and the final subtask consists of (4) publishing, which refers to presenting the results in a human-readable format. These four phases will be used as eligibility criteria in our systematic review.

To monitor the volume and diversity of available information, fact-checkers often rely on technologies such as automatic speech recognition, news alerts, and translation tools, which typically depend on underlying AI technologies. Since fake news spreads faster than real news, technologies could allow fact-checking organizations to be quicker and provide more far-reaching coverage than manual checking (Vosoughi, Roy, and Aral 2018). These authors also point out that automated claims verification seems like AI's ultimate application to fact-checking because many claims are not simply 'false' or 'true' but 'partially false' without additional context. One key role of professional fact-checkers is to help their audience understand the issue, despite showing a binary label. In this study, we will conduct a systematic analysis to check the scientific research published during the pandemic and understand how these concepts are documented.

To obtain data and identify a research focus, we will conduct a systematic review to help us find the most effective approach to contribute to the vast topic under investigation. We will adopt a systematic literature review (SLR) following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al. 2009). The guidelines suggest a checklist of 27 items under four phases in the systematic review process. We used four sources (ACM Digital Library, Scopus, Google Scholar, and Web of Science) to extract the most relevant literature. The search strategy revolves around the key-

words and search strings: (“Disinformation” OR “Health” OR “Pandemic”) AND (“Automat*” OR “Fact-checking”) AND (“journalism”) AND (“Informed citizens”). It was applied to capture the studies published during the Covid-19 pandemic from 2022 to 2020. The planning was done to discover potential contributions to journalism automation focused on addressing the health infodemic.

The search considered the following eligibility criteria—studies discussing fact-checking related to health and studies presenting technological automation solutions under the four phases of fact-checking automation (monitoring, spotting, checking, and publishing). To narrow the search, we applied the exclusion criteria of studies outside the Covid-19 years—2020 to 2022—and strictly theoretical research. The data collection process was planned to use the PICOT criteria, a convention from the Prisma methodology. The requirements are as follows: a) Population: pandemic, disinformation, health infodemic; b) Intervention: automatic fact-checking; c) Comparison: traditional journalism; d) Outcome: informed citizens; and e) Context: covid-19 pandemic, vaccination hoaxes. Seventy records were considered for full-text reading, and ten were excluded because they were strictly theoretical. Finally, 59 papers were selected to provide an overview of the automation technologies developed and used for fact-checking during the pandemic. This classification intends to generate insights for future approaches, given that we have no purpose in studying the technical details. Most of the studies focus on checking as the object of technological research (n=25), followed by spotting (20) and monitoring (18). Few studies addressed publishing technologies or even the theoretical approach to it (6). The studies are fundamentally analytical (18) and explanatory (6). Few cases make a comparative or descriptive investigation (3). The leading information/data gathering technique used is experimentation (19) and the collection of documents or audio-visual/digital media materials (i.e., tweets, posts, YouTube videos, WhatsApp messages) (13).

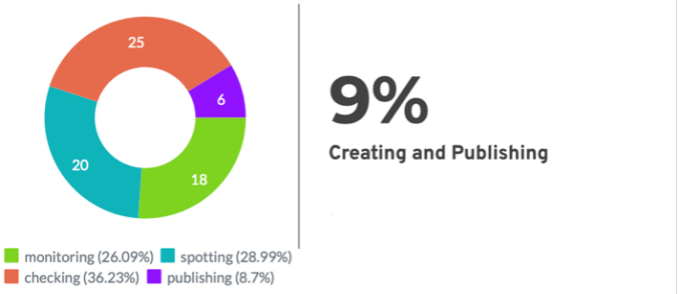

Figure 1: Less studied and reported according to the review.

We noticed only six publications in the “publication” category, representing 9% of the total (fig. 1). This finding reveals a relevant statistic showing an area that needs a better understanding of why it is less explored and the potential contributions that can be made through further research. Vedula and Parthasarathy (2021) introduce FACEKEG, an application that automatically performs explainable fact-checking. They present an output fact or claim and a model that constructs a knowledge graph. Gruzd and Mai (2020) present a case study of a hashtag used to propagate a false conspiracy theory and, through automation, verify the effectiveness of this case’s disclosure content, showing that the fake content was more efficient than fact-checking publishing content. Babakar Moy (2016) provides a theoretical review and Full Fact’s reports about the four phases in which fact-checking operates, including the “publication” phase as one of the highlights of automation. Ng, Cruickshank, and Carley (2022) analyzed fact-checked publications from Poynter and PolitiFact and further classified them into clusters by proposing a unique automated method, which can be used to classify diverse story sources in both fact-checked stories and tweets. In this “publication” category, we also include the following two articles that describe tagging fact-checking results with signaling, even though they do not directly feature NLP or NLG techniques. Lanius, Weber, and MacKenzie Jr (2021) present how under the COVID-19 infodemic, a flagging system could be built into automated fact-checking systems and other misinformation strategies. Likewise, Moravec, Kim, and Dennis (2020) also present a fake-news flag solution using two theoretical routes to deliver fact-checking results through automatic cognition and deliberate cognition.

These papers shed light on the potential of automation in the publication phase. They show that AI could be better explored to overcome limitations in advertising campaigns and strategies for sharing fact-checking results, as publishing automation tools and content propagation remain areas for further exploration. These findings indicate that multidisciplinary projects on solutions that help address the very problem that fake news tends to spread faster than verified news.

This bibliographical analysis shows that during the pandemic period, there were exciting and challenging possibilities, starting with the availability of technologies to improve the accuracy and speed of delivering verified news. Regarding the content reviewed, we propose some ideas to be explored considering certain limitations, especially concerning advertising campaigns and strategies to share, motivate, and reach the audience. As shown in the graph, a much smaller number of initiatives (9%) propose solutions around publishing automation and content propagation. This indicates that this topic has the potential for further research. It is worth noting that even within this small percentage, no solutions were found using NLP, synthetic media, automatic sharing, or chatbot interfaces.

Funding

This research is financed with national funds through the Portuguese Funding Agency FCT, within the grant 2021.05707.BD.

References

Arnold, P. 2020. The challenges of online fact checking. Technical report, Technical report, Full Fact.

Babakar, M.; and Moy, W. 2016. The State of automated fact-checking. Full Fact.

Gruzd, A.; and Mai, P. 2020. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society*, 7(2): 2053951720938405.

Lanius, C.; Weber, R.; and MacKenzie Jr, W. I. 2021. Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey. *Social Network Analysis and Mining*, 11(1): 32.

Miller, I. 2020. Research Guides: Journalism: Fact-Checking Sites. *Eastern Washington University. Retrieved*, 17.

Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D. G.; and PRISMA Group*, t. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4): 264–269.

Moravec, P. L.; Kim, A.; and Dennis, A. R. 2020. Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research*, 31(3): 987–1006.

Nakov, P.; Corney, D.; Hasanain, M.; Alam, F.; Elsayed, T.; Barrón-Cedeño, A.; Papotti, P.; Shaar, S.; and Martino, G. D. S. 2021. Automated fact-checking for assisting human fact-checkers. *arXiv preprint arXiv:2103.07769*.

Ng, L. H. X.; Cruickshank, I. J.; and Carley, K. M. 2022. Coordinating Narratives Framework for cross-platform analysis in the 2021 US Capitol riots. *Computational and Mathematical Organization Theory*, 1–17.

Vedula, N.; and Parthasarathy, S. 2021. Face-keg: Fact checking explained using knowledge graphs. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 526–534.

Vlachos, A.; and Riedel, S. 2014. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, 18–22.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *science*, 359(6380): 1146–1151.