# Psychological interventions countering misinformation in social media. A scoping review. Research protocol

**Authors:**

Paweł Gwiaździński 1,2, Jonas Kunst 3, Aleksander B. Gundersen 3, Karolina Noworyta 4, Agata Olejniuk 5, Jan Piasecki 1

**Contributors:**

Ositadima Chukwu6, Łucja Zaborowska6

**Affiliations:**

1. Department of Philosophy and Bioethics, Faculty of Health Sciences, Jagiellonian University Medical College, Kopernika 40, 31-501 Kraków, Poland
2. Consciousness Lab, Institute of Psychology, Jagiellonian University, Ingardena 6, 30-060 Kraków, Poland
3. Department of Psychology, University of Oslo, Postboks 1094 Blindern, 0317 Oslo, Norway
4. Affective Cognitive Neuroscience Laboratory, Department of Pharmacology, Maj Institute of Pharmacology of the Polish Academy of Sciences, Smętna 12, 31-343 Kraków, Poland
5. Poznan University of Technology, Plac Marii Skłodowskiej-Curie 5, 60-965 Poznań
6. Faculty of Medicine, Jagiellonian University Medical College, Anny 12, 31-008 Kraków, Poland

**Conflicts of interest:**

The authors do not declare any conflicts of interest.

**Abstract**

*Introduction*

Misinformation is a complex concept and its meaning can encompass several kinds of different phenomena. Liang Wu et el. consider a wide variety of online behavior as misinformation:[1] unintentionally spreading false information, intentionally spreading false information, disseminating urban legends, sharing fake news, unverified information, and rumors, as well as crowdturfing, spamming, trolling, and propagating hate speech, or being involved in cyberbullying. The aim of this review is to address the following question: "What psychological interventions countering misinformation can be deployed on popular social media platforms (e.g. Twitter, Facebook)?". In order to address this question, we have designed a systematic scoping review procedure in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[2,3]

Effective measures of countering misinformation on social media are instrumental for facilitating and fostering reliable public conversation about political and social problems. Moreover, countering misinformation on social

media platforms can be also considered a public health intervention, especially in the time of health emergencies, such as the COVID-19 pandemic.

*Methods and analysis*
A scoping review is a modern, rigorous approach to synthesis science, developed among others by the Joanna Briggs Institute team. For data extractions, we plan to use the following databases: Embase, Scopus, and PubMed. For paper selection, eligibility criteria were defined.

**Keywords**: misinformation, fake news, disinformation, social media, psychological intervention, fact-checking, red-flagging, media literacy, scoping review, systematic review, scoping review protocol

**Introduction**
Misinformation is a complex concept that can encompass several kinds of different online behaviors. The spread of misinformation is facilitated by the growth of social media platforms that turn a regular user not only into a consumer of information but also its producer and carrier. Therefore, individuals can produce and spread information on an unprecedented scale. However, it is not always clear what the term "misinformation" denotes, and what kinds of online behavior constitute spreading misinformation. For instance, Liang Wu et al. consider a wide variety of online behavior as a kind of misinformation:[1] unintentionally spreading false information, intentionally spreading false information, disseminating urban legends, sharing fake news, unverified information, and rumor, as well as crowdturfing, spamming, trolling, and propagating hate speech, or being involved in cyberbullying. Other authors define misinformation in the following way: "[Misinformation] is wrong or false information circulated as a result of a genuine mistake, omission, prejudice or sheer ignorance";[4] "Misinformation is unintentional dissemination of false information";[5] "Misinformation refers to the inadvertent sharing of false information";[6] "Misinformation is inaccurate information that is the result of an honest mistake or of negligence";[7] "Misleading or inaccurate information shared by people who do not recognize it as such".[8] Whereas scholars have more and more understanding of what misinformation is and how to recognize it, it is still very unclear how to counter its effects and spread, especially in social media.

This systematic scoping review protocol is prepared in accordance with the PRISMA guidelines and will be pre-registered in a publicly available database for systematic review protocols. The main goal of the scoping review is to give a systematic overview of the available psychological interventions countering misinformation in social media. We define social media as:

> ... any Web-based computer-mediated tools to cocreate, share or exchange information, ideas, pictures, or videos in virtual communities and networks (such as message boards, social networks, patient forums, Twitter, blogs, and Facebook).[9]

Research has shown that 59% of Twitter articles are never opened and read in detail by sharers.[10] This finding indicates that most of the shared content avoids scrutiny and critical evaluation by users, and is propagated based on emotional responses rather than on rational grounds.[11] Another important issue is highlighted by the rising amount of evidence for the radicalizing effect of YouTube algorithms.[12] This effect might also be present in other social media.

The spread of misinformation can also cause public security threats as the so-called pizzagate incident has shown.[13] In the case of pizzagate, a group of 4chan conspiracy theorists motivated a man to attack a pizza restaurant in Washington, DC, in an act of domestic terrorism.

On a global scale, misinformation can pose a danger to liberal democracy as several studies have claimed.[11, 14] Also, public health is endangered by the spread of misinformation. For instance, in the times of the global COVID-19 pandemic, misinformed beliefs cause reluctance towards vaccinations in the population,[15] which is rooted in misinformation about vaccines in general.[16,17] Moreover, misinformation can influence people to try untested drugs or dangerous therapies, which sometimes results in death,[18] and to refuse to adhere to safety measures.[19] Finally, misinformation itself can be thought of as a kind of epidemic or an infodemic. In the times of the global pandemic crisis, some scholars have outlined similarities between the spread of the virus and the spread of the misinformation about it.[19, 20] This second kind of epidemic might be as dangerous as the virus itself. This view urges the question of whether it is possible to increase the immunity to misinformation online (here referred to as "webimmunity") through interventions that reduce the spread of misinformation from user to user. The growing amount of medical misinformation poses a real threat to public health and safety. Conspiracy theories, including medical conspiracies,[21] lead people to try dangerous treatment methods, while refraining from seeking professional medical help. For instance, research has shown that the risk of mortality for people who use alternative cancer therapies has an overall hazard ratio, after a 5-year period, as high as 2.5. In other words, on average, choosing alternative medicines alone more than doubled the risk of death.[22] This shows that misinformation can be a life-threatening danger and highlights how dire this problem may become.

In this systematic scoping review, we aim to address this particular issue by answering the following research question: "What psychological interventions countering misinformation can be deployed on social media platforms?".

The aim of scoping reviews is to synthesize and map existing data related to defined research questions. Scoping reviews do not appraise reviewed evidence, as their main goal is to synthesize state-of-the-art data in a transparent and systematic manner.

**Methods**

The scoping review method is a modern, rigorous approach to synthesis science, developed by the JBI team.[23] The objectives, inclusion criteria, and methods for this scoping review were specified in advance and documented in a protocol. For data extractions, we plan to use the following databases: Embase, Scopus, and PubMed. We will extract and summarize details of the eligible studies using a data extraction document, prepared prior to search attempts. Two review authors (1 and 6) will extract data independently and will resolve disagreements by discussion. Where data are missing from reports, we will attempt to contact the study authors to obtain this information. When a study meets the eligibility criteria, we will extract only data about intervention outcomes, including control and study group comparisons.

**Inclusion criteria**

Our inclusion and exclusion criteria were carefully defined to encompass as wide a range of papers as possible, while not including works that are non-informative for the assumed research questions and scoping review methods. We were inspired by both the classical approach to systematic reviews, and by more modern approaches, focused on qualitative methods of reviews.[24]

- A paper, to be included in the sample, has to meet all of the criteria listed below:
- The paper broadly focuses on one form of misinformation (i.e., spread of false information, urban legends, fake news, fake science, etc.).
- The paper addresses the issues of misinformation in social media (for instance: Facebook, Twitter, Instagram, etc., or pairings of those).
- The paper is empirical, that is, present primary data obtained through a qualitative and/or quantitative research methodology.
- The paper has been published since 2004 (the year Facebook was created – the oldest social network of the modern scale).
- The paper proposes a psychological intervention to counter misinformation and test it. Psychological intervention is understood here as an intervention or experimental manipulation that targets psychological, intermediary, or cognitive processes, or actual behavior. An example of psychological intervention might be asking subjects to pause to consider why a headline is true or false, before sharing.[25] An intervention is not psychological when it targets, e.g., either biochemical functions of a body (e.g. pharmacological intervention) or the functions of a computer/phone (e.g. computer processing information on a phone). A compatible definition of the intervention considered in this review is the one that can be found in the APA Dictionary of Psychology: 'strategies and processes designed to measure the change in a situation or individual after a systematic modification (diet, therapeutic technique, etc.) has been imposed or to measure the effects of one type of intervention program as compared to those of another program. Experimentation is the most common type of intervention research, but clinical trials and qualitative studies may also be used'.[26]
- As such, experimental manipulations aimed at reducing susceptibility to misinformation in social media will be included in this review.
- The paper is a peer-reviewed research paper published in an international journal, etc.
- The paper has been published in English.

Exclusion criteria will be the following:
- Reviews, meta-analyses, theoretical, or other non-empirical papers.
- Non-English texts.
- Papers not addressing misinformation, urban legends, fake news, fake science, etc. in the context of social media.
- Papers testing and/or considering interventions which are speculative or impossible to employ in a social media environment: for instance, interventions requiring the involvement of highly trained specialists are not viable interventions for the purpose of this review.

**Data synthesis**
For data synthesis, we decided to employ a qualitative approach and descriptive statistics. We plan to combine the details of included studies in a narrative review. We will take into account the methodological heterogeneity, the ways misinformation is operationalized and the type of intervention (for instance: intervention on the user side and intervention on the software side should be considered separately), duration (the time it takes for an intervention to be successful), effortfulness and/or difficulty of application, and eventual follow-up study outcomes (to establish whether an intervention has left persisting effects among users). We will also collect information about the journal of publication, year of publication, the country of a study, the sponsor of the study (university, public funds, industry).

In the preliminary phase, A, B (under 1 supervision), 1, and 2 screened first 100 records for eligibility criteria. The goal of the preliminary phase was to assure uniform understanding of eligibility criteria.

**Risk of bias assessment**

Two independent reviewers will perform searches, review the papers, and extract data. We will be using third-party developed software to compile and structure our database. Discrepancies in data synthesis will be discussed and resolved, with input from a third-party jury if needed.

**Search strategy**

In terms of the search strategy, we followed the JBI recommendations[23]: in the first phase, we search databases for eligible papers in Scopus, PubMed, and Google Scholars using a loose set of keywords ("misinformation", "fake news"). Next, we analyzed the terminology used in eligible papers, we have chosen and designed a comprehensive searching string – inspired by the PICO strategy. The resulting searching string will be applied in Embase, Scopus and PubMed. Then in the third phase, we will search references of all articles meeting inclusion criteria for additional sources.

| P – population | ("disinformation" OR "misinformation" OR "fake news" OR "conspiracy theor*" OR "urban legend*" OR "rumor*" OR "hate speech" OR "cyberbullying" OR "fake science" OR "mislead*" OR "fake source*" OR "propagand*") AND ("social media" OR "facebook" OR "instagram" OR "twitter" OR "tiktok" OR "youtube" OR "messenger" OR "whatsapp" OR "telegram" OR "internet" OR "media" OR "blog*" OR "reddit" OR "4chan") |
|---|---|
| I – intervention | ("intervent*" OR "tag*" OR "factcheck*" OR "false-tag" OR "refutation" OR "correct*" OR "retraction" OR "flag*" OR "headline*" OR "counter*" OR "rated false" OR "disrupted" OR "questionnaire*" OR "survey*" OR "interview*" OR "focus group*" OR "case stud*" OR "observ*" OR "experiment*" OR "qualitative" OR "quantitative" OR "mixed method*" OR "experiment*") |
| C – comparison | ("view*" OR "experienc*" OR "opinion*" OR "attitude*" OR "perce*" OR "belie*" OR "judge*" OR "feel*" OR "know*" OR "understand*" OR "assess*" OR "expect*" OR "tenden*") |
| O – outcome | ("share*" OR "verify" OR "follo*" OR "unfollo*" OR "subscrib*" OR "unsubscrib*" OR "click*" OR "induc*" OR "trust*" OR "distrust*" OR "check*" OR "reduc*" OR "judge*" OR "inferenc*" OR "correct*" OR "reflect*" OR "reliance" OR "resist*" OR "back-fire" OR "influe*" OR "like") |

((P) AND (I) AND (C) AND (O))

(("disinformation" OR "misinformation" OR "fake news" OR "conspiracy theor*" OR "urban legend*" OR "rumor*" OR "hate speech" OR "cyberbullying" OR "fake science" OR "mislead*" OR "fake source*" OR "propagand*") AND ("social media" OR "facebook" OR "instagram" OR "twitter" OR "tiktok" OR "youtube" OR "messenger" OR "whatsapp" OR "telegram" OR "internet" OR "media" OR "blog*" OR "reddit" OR "4chan")) AND ("intervent*" OR "tag*" OR "factcheck*" OR "false-tag" OR "refutation" OR "correct*" OR "retraction" OR "flag*" OR "headline*" OR "counter*" OR "rated false" OR "disrupted" OR "questionnaire*" OR "survey*" OR "interview*" OR "focus group*" OR "case stud*" OR "observ*" OR "experiment*" OR "qualitative" OR "quantitative" OR "mixed method*" OR "experiment*") AND ("view*" OR "experienc*" OR "opinion*" OR "attitude*" OR "perce*" OR "belie*" OR "judge*" OR "feel*" OR "know*" OR "understand*" OR "assess*" OR "expect*" OR "tenden*") AND ("share*" OR "verify" OR "follo*" OR "unfollo*" OR "subscrib*" OR "unsubscrib*" OR "click*" OR "induc*" OR "trust*" OR "distrust*" OR "check*" OR "reduc*" OR "judge*" OR "inferenc*" OR "correct*" OR "reflect*" OR "reliance" OR "resist*" OR "back-fire" OR "influe*" OR "like")

## Strategy for data extraction

Two review authors (1 and 6) will independently perform data extraction in accordance with the previously selected data extraction form. Any disagreements will be resolved by discussion, and, when necessary, a third person – an arbiter – will be involved. The data extraction form will contain the following information:

**Study Details**: location, year, contextual information.

**Methods**: study design, the total duration of the study, study location, study setting (online or in the lab), risk of bias information, withdrawals.

**Participants**: N, mean age or age range, inclusion criteria, exclusion criteria, gender distribution, and other demographic variables and characteristics.

**Intervention**: description of the intervention, control condition, and potential co-interventions, how viable the application of this intervention to social media is.

**Outcomes**: description of outcomes.

Other information: funding and notable conflicts of interest of authors.

Lastly, we will use Google Scholar as a supplementary tool to check the validity of the obtained results.

## Data storage

Each step of the scoping review will be documented. All notes and data will then be stored in secure One Drive cloud documents.

## Expected results

The expected outcomes of this study are:

A comprehensive review of psychological interventions aimed at countering misinformation and related phenomena in social media.

Effectiveness and persistence of different kinds of psychological interventions.

Viability of reviewed interventions to be effective in social dimension (so-called webimmunization).

## Type and method of review

JBI Scoping systematic review.

Anticipated or actual start date: 28.04.2021

Anticipated completion date: 30.09.2021

**Stage of the review at the time of this submission**

| Stage | Started | Completed |
|---|---|---|
| Preliminary searches | Yes | Yes |
| Piloting of the study selection process | Yes | Yes |
| Formal screening of search results against eligibility criteria | Yes | No |
| Data extraction | No | No |
| Risk of bias (quality) assessment | Yes | No |
| Data analysis | No | No |

**References:**
1. Wu L, Morstatter F, Carley KM, Liu H. Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter* 2019; Nov 26;21(2):80-90. Doi: 10.1145/3373464.3373475.
2. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 Statement. *Systematic reviews* 2015; Dec 4(1):1-9. Doi: 10.1186/2046-4053-4-1.
3. Peters MD, Godfrey C, McInerney P, Baldini Soares C, Khalil H, Parker D. Chapter 11: scoping reviews. *Joanna Briggs Institute Reviewer's Manual* 2017. doi:10.46658/jbirm-20-01.
4. Bittman L. The use of disinformation by democracies. *International Journal of Intelligence and Counter Intelligence* 1990; Jan 1;4(2):243-61.
5. Kragh M, Åsberg S. Russia's strategy for influence through public diplomacy and active measures: the Swedish case. *Journal of Strategic Studies* 2017; Sep 19;40(6):773-816.
6. Wardle C. Fake news. It's complicated. *First Draft* 2017; Feb 16;16:1-1.
7. Haiden, L., McManus, C., Michaud, C., Duffy, E., Ranautta-Sambhi, K. and Ilbury, M., 2018. Fake News. A Roadmap. NATO Strategic Communication. https://issuu.com/natostratcomcoe/docs/fake_news_book_final_full_version__
8. de Cock Buning M. A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation.2018: https://publications.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d01aa75ed71a1/language-en [Accessed 30 Jul. 2018].

9.  Obar JA, Wildman SS. Social media definition and the governance challenge-an introduction to the special issue. Obar, JA and Wildman, S.(2015). *Social media definition and the governance challenge: An introduction to the special issue. Telecommunications policy* 2015; Jul 22;39(9):745-50. 10.2139/ssrn.2637879.

10. Gabielkov M, Ramachandran A, Chaintreau A, Legout A. Social clicks: What and who gets read on Twitter?. *Proceedings of the 2016 ACM SIGMETRICS International Conference On Measurement and Modeling of Computer Science* 2016; Jun 14: 179-192. Doi: 10.1145/2964791.2901462.

11. Boler M, Davis E. *Affective Politics of Digital Media: Propaganda by Other Means*. 2020. doi:10.4324/9781003052272

12. Ribeiro MH, Ottoni R, West R, Almeida VA, Meira Jr W. Auditing radicalization pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* 2020; Jan 27:131-141.

13. Pizzagate conspiracy theory 2021. Wikipedia. https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory

14. Jolls T, Johnsen M. Media literacy: a foundational skill for democracy in the 21st century. *Hastings LJ* 2017;69:1379.

15. Burki T. Vaccine misinformation and social media. *The Lancet Digital Health* 2019; Oct 1;1(6):e258-9.

16. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics* 2020; Jul 11:1-8.

17. Haghdoost Y. Alcohol poisoning kills 100 Iranians seeking virus protection. Bloomberg. com 2020. https://www.bloomberg.com/news/articles/2020-03-18/alcohol-poisoning-kills-100-iraniansseeking-virus-protection. Accessed: 2020-04-10.

18. Bierwiaczonek K, Kunst JR, Pich O. Belief in COVID-19 conspiracy theories reduces social distancing over time. *Applied Psychology: Health and Well-Being* 2020; Dec;12(4):1270-85.

19. Call for Action: Managing the Infodemic: https://www.who.int/news/item/11-12-2020-call-for-action-managing-the-infodemic. Accessed: 2021-02-28.

20. Zarocostas J. How to fight an infodemic. *The Lancet* 2020; Feb 29;395(10225):676. Doi: 10.1016/S0140-6736(20)30461-X.

21. Andrade G. Medical conspiracy theories: cognitive science and implications for ethics. *Medicine, Health Care and Philosophy* 2020; Apr 16:1-4. doi:10.1007/s11019-020-09951-6.

22. Johnson SB, Park HS, Gross CP, Yu JB. Use of alternative medicine for cancer and its impact on survival. *JNCI: Journal of the National Cancer Institute* 2018; Jan;110(1):121-4.

23. Peters M, Godfrey C, McInerney P, Soares C, Khalil H, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews.

24. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative health research* 2012; Oct;22(10):1435-43. Doi: 10.1177/1049732312452938.

25. Fazio L. Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review* 2020 Feb 10;1(2). 10.37016/mr-2020-009.
26. Van den Bos GR. *APA dictionary of psychology*. American Psychological Association; 2007. https://dictionary.apa.org/intervention-research.