Online Journal of Communication and Media Technologies, 2022, **12**(3), e202217
e-ISSN: 1986-3497

OPEN ACCESS

Research Article

Developing Fake News Immunity: Fallacies as Misinformation Triggers During the Pandemic

Elena Musi [1*]
0000-0003-2431-455X

Myrto Aloumpi [2]
0000-0003-2636-2088

Elinor Carmi [3]
0000-0003-1108-2075

Simeon Yates [1]
0000-0002-7298-8826

Kay O'Halloran [1]
0000-0002-7950-0889

[1] Department of Communication and Media, University of Liverpool, Liverpool, UK
[2] Department of Philology, University of Crete, GREECE
[3] Department of Sociology, City University London, London, UK
[*] Corresponding author: elena.musi@liverpool.ac.uk

Citation: Musi, E., Aloumpi, M., Carmi, E., Yates, S., & O'Halloran, K. (2022). Developing Fake News Immunity: Fallacies as Misinformation Triggers During the Pandemic. *Online Journal of Communication and Media Technologies, 12*(3), e202217. https://doi.org/10.30935/ojcmt/12083

ARTICLE INFO

Received: 17 Nov 2021

Accepted: 4 May 2022

ABSTRACT

Misinformation constitutes one of the main challenges to counter the infodemic: misleading news, even if not blatantly false, can cause harm especially in crisis scenarios such as the pandemic. Due to the fast proliferation of information across digital media, human fact-checkers struggle to keep up with fake news, while automatic fact-checkers are not able to identify the grey area of misinformation. We, thus, propose to reverse engineer the manipulation of information offering citizens the means to become their own fact-checkers through digital literacy and critical thinking. Through a corpus analysis of fact-checked news about COVID-19, we identify 10 fallacies–arguments which seem valid but are not–that systematically trigger misinformation and offer a systematic procedure to identify them. Next to fallacies, we examine the types of sources associated to (mis-/dis-)information in our dataset as well as the type of claims making up the headlines. The statistical patterns surfaced from these three levels of analysis reveal a misinformation ecosystem where no source type is exempt from flawed arguments with frequent evading the burden of proof and cherry picking behaviors, even when descriptive claims are at stake. In such a scenario, exercising the audience's critical skills through fallacy and semantic analysis is necessary to guarantee fake news immunity.

Keywords: misinformation, fallacy theory, digital literacy, fact-checking, multi-level annotation

INTRODUCTION

One of the major challenges of the current information ecosystem is the rapid spread of misinformation through digital media. Initial discussions of infodemiology–the role of information spread in support of or exacerbating issue of health and health policy–has brought to the fore the need to improve fact-checking to counter intentional and unintentional misbehaviors and inform policy making. The buzzword "fake news" has been used to refer to phenomena ranging from news, parody, to propaganda, and news fabrication. Even when adopting a strict definition of fake news as intentionally lacking facticity to a certain degree (Tandoc et al., 2018), there are clear variations: a news claiming that "mRNA vaccines are capable of altering or damaging human DNA" (Kasprak, 2020) is more fake than a news claiming that "vaccines are unavoidably unsafe" (Teoh, 2020b). Both might trigger wrong perceptions and attitudes, but the latter news claim does not convey entirely false information. As explained by the fact-checker organization *Healthfeedback.org*, the legal phrase "unavoidably unsafe", which takes into account risk/benefit trade-offs, leads to misleading interpretations of the vaccine as "dangerous".

Due to continuous updates about COVID-19 from the scientific community as well as governments and health institutions, the media may unintentionally disseminate misleading content which goes beyond lexical vagueness by, for example, drawing defeasible generalizations out of partial scientific results or single anecdotes (Archila et al., 2019). In other words, what makes these types of news *fake* is not just the truth of the information conveyed. Rather it is the misleading presentation or reasoning of the arguments they convey. This is done, for example, through false analogies, hasty generalizations, and cherry picking of information. This type of fake news is generally addressed as *misinformation,* which is the distribution of information which is not necessarily false and not deliberately created to harm (Yates et al., 2020). Even though unintentionally dangerous, misinformation has a wide societal impact. Brennen et al. (2020) found that 59% of fake news does not contain either fabricated or imposter content, but rather reconfigured misinformation. This misinformation proliferates through social media, the main source of news for infodemically vulnerable citizens. In other research (Carmi et al., 2020), limited types of source and information checking across both social media and search engine and reliance on the opinions of close friends and family have been identified as corresponding to low levels of digital and data literacy.

However, the identification of misinformation is far from being successfully addressed by human fact-checkers, let alone automated ones. The rating categories of different fact-checker organizations represented in the Google Fact Checker initiative lack of an agreed truth barometer based on systematic, mutually exclusive and clear criteria, thus hindering public understanding. As a result, datasets coded as misinformation that can be used to train systems for automatic fact-checking of information are scarce, even though needed. As remarked by Thorne and Vlachos (2018), current text classification approaches leveraging fact-checked datasets of claims are not enough since additional contextual information alongside factuality is required to capture misinformation.

To lay the foundations for a fact-checking process that uncovers misinformation triggers, we propose a systematic and multilevel procedure to identify fallacious arguments. Our theoretical assumption is that fallacies, arguments that seem valid but are not, work as indicators of misinformation. We apply our system to the analysis of a dataset of 1,135 COVID-19 related fact-checked news, revealing major trends in the way misinformation is constructed and communicated. From an empirical perspective, we adopt a bottom-up approach focusing on the specific characteristics of the news reports: that is, we develop a set of guidelines for the identification of:

a. fallacies (e.g., false authority);
b. the type of media source hosting the news (e.g., social media; broadcast digital news);
c. the semantic type of claim expressing the news title (e.g., prediction vs. interpretation).

We conduct an annotation experiment with two non-expert annotators and then check disagreement cases emerging from the inter-annotator agreement metrics through the aid of an expert annotator (golden standard annotation). We then focus on statistical trends which feature the golden standard annotations looking at the frequency of values for each analytic level as well as $\chi 2$ contingency tables across different levels of analysis to answer the following questions:

1. **RQ1:** Is it possible to develop a reliable procedure for the identification of misinformation triggers?
2. **RQ2:** What are the triggers of misinformation (fallacies, types of claims) and what is their frequency?
3. **RQ3:** What sources are more likely to spread misinformation?
4. **RQ4:** Do certain sources tend to be associated with certain fallacies and/or types of claims and vice versa?

In the following sections, we show how we developed our theoretical approach based on fallacy theory (Hamblin, 2022). We then move onto explaining how we designed the classification system of most common fallacies in news relating to COVID-19. After that, we zoom into the categories that have triggered agreement and disagreement among the annotators. We then move to the results of the analysis we conducted of the news articles pointing to statistically significant trends. Finally, we discuss our findings and how they can contribute to educate society about online news manipulations.

THEORETICAL FRAMEWORK

Digital Media Literacy to Fight the Infodemic

The COVID-19 pandemic meant that millions of people across the world were moving in and out of lockdowns and had to rely on digital systems and news sites for their everyday needs. But beyond digital divides around access to the Internet there is also the issue of digital media literacy. For example, Abdulai et al. (2020) have examined COVID-19 related digital skills among people in Ghana and argue that people experienced challenges in locating the appropriate online resources related to the pandemic. Importantly, they found that people experienced difficulties in distinguishing good quality information from opinions and anecdotes. Similarly, Beaunoyer et al. (2020) argue that people who have lower digital health related skills are more vulnerable to getting infected and infecting others because they have more challenges in accessing, understanding and applying the proper measures. As they argue, "people not able to decipher the degree of veracity of information (typically due to low level of critical digital or health literacy) might follow various advice regarding COVID-19 that could not only be detrimental for their health but also be harmful for the population" (Beaunoyer et al., 2020). One of the avenues they propose to mitigate digital inequalities related to COVID-19 is to improve people's ability to detect fake-news.

According to Fletcher et al. (2020), in the UK there was an interest in news in the beginning of the pandemic that slowly decreased. However, access to news about COVID-19 was unevenly distributed, with people who come from lower socio-economic status in terms of levels of education (this factor is especially dominant in online news consumption) and household income being less likely to consume news. As Fletcher et al. (2020) identified throughout the pandemic, people used social media in high proportions but as time progressed the use of social media for news and information about COVID-19 decreased. Nevertheless, the proportion of people who say they avoid news increased to 25% in early June 2020, a trend that is influenced by various factors such as the negative effect on mood.

In relation to engagement with fake news and misinformation, Kyriakidou et al. (2020) argue that UK citizens "felt misled by a range of information they encountered, which–in their view–was often conflicted or inconclusive, including government claims about the human impact of the pandemic in the UK". According to them, people felt that the most confusing and misleading content they encountered came from the UK government's messages during the pandemic. In this context, some scholars (Amazeen & Bucy, 2019; Kahne & Bowyer, 2017; Vraga et al., 2020) argue that teaching people news literacy might be one solution. News literacy is defined as having an understanding about the processes of producing, distributing, and engaging with news. More specifically, news literacy can "provide a foundation to improve information consumption processes by giving social media users the tools to identify, consume, and share high-quality information regarding COVID-19" (Vraga et al., 2020).

In the era of networked society, to be able to responsibly consume and produce news implies being a media literate person who "can decode, evaluate, analyze and produce both print and electronic media. The fundamental objective of media literacy is a critical autonomy relationship to all media" (Aufderheide, 1993). The centrality of media literacy to counter fake news has been recently underlined by the European Commission in their action plan against disinformation as requiring "continuous and sustained efforts to support education and media literacy, journalism, fact-checkers, researchers, and the civil society as a whole" (2018)[1]. Scholars have repeatedly pointed to critical thinking as the kernel of media literacy. Hobbs (2011) considers, for instance, "comprehending messages and using critical thinking to analyze message quality, veracity, credibility […]" as the second component of the five essential (access; analyze and evaluate; create;

[1] Action plan against disinformation. "Joint Communication the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions" (2018). https://ec.europa.eu/newsroom/dae/document.cfm

reflect; act) to develop media literacy. Similarly, Koltay (2011) defines "having a critical approach to quality and accuracy of content" among the five stages to build media literacy. However, so far, no systematic intervention to teach critical thinking in the news environment has been carried out. The ability to evaluate whether the arguments that form news are correct or fallacious contributes to this endeavor, constituting part and parcel of the critical thinking needed to be a digital media literate.

Rhetorical Clues (Fallacies) To Identify Misinformation

The theoretical basis of our approach is founded on the notion of fallacy. A standard definition of fallacy that goes back to Aristotle is an argument that "seems to be valid but is not so" (Hamblin, 2022; Tindale, 2007). Aristotle has undoubtedly provided the foundations for the systematic study of fallacious arguments[2], even if the textbook versions he neatly outlines may be rarely found in real life discourse. Because fallacious arguments can be very close to valid ones sometimes it may be difficult to talk about clear-cut distinctions (Boudry et al., 2015). More significantly, this closeness explains why fallacious reasoning is persuasive because it follows, even if partially, the patterns and tropes of non-fallacious reasoning thereby producing arguments that are not entirely invalid or outrageously unacceptable, at least at first glance.

With a focus to the realm of mis-/dis-/information, the persuasiveness of misinformation can be explained in a similar vein: fake news can be viewed as news that 'seems to be valid but is not so'. For example, the fallacy of 'cherry-picking' may happen intentionally or unintentionally when specific information that supports a given position is chosen, while ignoring or dismissing information which does not support it[3]. This means that an instance of fake news that is the outcome of cherry-picking can be based on *partial* information, but not necessarily *false* information. Such combinations of valid and invalid information, and arguments that we often encounter in discourse that involves fallacious reasoning, shows why misinformation has a grip on people. This becomes even more evident when we turn to news in the realm of misinformation. Fallacy identification is an efficient way for achieving bottom-up deconstruction of misinformation that privileges misinformation pre-bunking over debunking.

While fact-checking websites attempt to categorize misinformation on the basis of truth barometers that are partially informative (e.g., labels such as "half true"), fallacy identification points directly to the roots of the misinformation problem. In particular, fallacy identification copes with the grey areas of misinformation and allows us to draw and analyze its different shades in a qualitative and constructive way that could never be achieved through the available truth barometers. Importantly, it helps us learn how to identify misinformation and cope with online manipulations.

The relevance of fallacies can be showcased through the analysis of news from our COVID-19 dataset. A claim circulated on Facebook that 'the flu shot causes false positive results on COVID-19 tests' has been fact-checked by *Healthfeedback.org* and assigned the label "incorrect". How helpful, however, is the label "incorrect" or, to take a few more from the same truth barometer, "misleading", "half true", and "inaccurate" for evaluating and deconstructing misinformation? Such labels merely indicate that there is something flawed with the news at hand, but they do not provide constructive insights about the nature of misinformation.

Fallacy identification, on the other hand, explains the roots of misinformation, whether it relies on quantity and quality of evidence available, the type of reasoning at stake or the language involved. In the case at hand, for example, the dominant fallacy is that of *post hoc*: the fact that coronavirus was detected in some individuals who received the flu shot does not prove that the flu shot caused the detection[4]. If the label "incorrect" warns us that there is something problematic with a piece of news, the label *post hoc* takes us several levels deeper by allowing us to identify the level (reasoning) and the origin of misinformation.

---

[2] *Sophistical Refutations; Rhetoric 2.24.* The first theoretical discussions of fallacious reasoning can be traced back to Gorgias (now mostly lost and fragmented) and Plato (e.g., Hippias Minor, Euthydemus).

[3] One of the earliest acknowledgements of cherry-picking appears in Plato *Hippias Minor* 369bc.

[4] This may be picked up in the detailed explanation provided by the fact-checkers but is not reflected in their labelling system. See https://healthfeedback.org/claimreview/claim-that-flu-shot-causes-false-positive-results-on-Covid-19-tests-is-unsupported-and-misleading/

Fallacies also take us in a new direction when observing and understanding broader trends in misinformation. The taxonomy of ten fallacies that we employ, which is based on Tindale's (2007) framework, falls under four broader classes:

a. fallacies from diversion, that divert the attention from the real issue at hand;

b. structural, linked to the quantity of arguments;

c. logical fallacies;

d. language fallacies.

This broader categorization enables us to understand patterns in the spread of misinformation. For example, there seems to be a correlation between news based on the use of images and videos, and fallacies from diversion, especially "red herring" (the arguments are not relevant for the conclusion) and "strawman" (when the other side's arguments are intentionally misrepresented). In such cases, images or videos are taken out of their original context and are employed as evidence for unrelated stories.

An instance of such misinformation is re-labelling images of crowds in demonstration as evidence for people rising up for COVID-19 related issues, whereas in fact those images are taken out of their original context which has nothing to do with COVID-19 related demonstrations. Going beyond the analysis and deconstruction of specific cases of misinformation, fallacy classes allow us to identify and understand patterns in the spread of misinformation that can be peculiar to specific media and types of news.

Data

Our data comprises all the COVID-19 news that have been fact-checked by the five fact-checkers in English: *Snopes.com*; *Healthfeedback.org*; *Politifact.com*; *Fullfact.org*; and *TheFerret.scot*. Our timeframe for the data collection is from the beginning of the outbreak in January 2020 till end of June 2020, where we collected 1,135 news articles. We have webcrawled the fact-checkers' official sites and created a dataset that contains the following information: fact-checked news claim, link to the full fact-checked news, fact-checkers' comments, and fact-checkers' ratings.

It is important to emphasize that the various fact-checkers have different truth barometers in terms of number and categories of ratings: while, for example, *TheFerret.scot*. uses a scale of seven ratings pointing to different degrees of veridicality (e.g., "mostly true", "false"), *Snopes* adopts a list of 14 ratings ranging from "mixture" to "misattribution" or "scam". Despite, such variations, all the ratings allow to disentangle information deemed as reliable (true information) information which constitutes complete fakery (disinformation) and information which contains elements both of truth and of falsity (misinformation). Since *Fullfact.org* does not have a set of fixed ratings tagged onto the fact-checked news, each fact-checked news article has been manually analyzed (**Table 1**). In our dataset, disinformation constitutes 44% of the fact-checked news, true information amounts to 9% and misinformation covers 46% of the cases, confirming that misleading news form a consistent portion of news flagged as fake.

Table 1. Dataset of COVID-19 fact-checked news

Fact-checker	Disinformation news	Information news	Misinformation news	Total fact-checked news
Snopes	94	46	80	220
Health Feedback	2	0	68	70
The Ferret	27	0	13	40
Full Fact	46	31	208	285
Politifact	335	28	157	520
Total	504	105	526	1,135

The review of the descriptions of the fact-checking processes disclosed by the fact-checkers reveals that there are no common procedures for identifying which news to fact-check. However, we did identify several common factors which influence the decision to choose news articles. These include:

a. newsworthiness;

b. popularity across media;

c. potential harm.

As a result, our dataset of fact-checked news is not balanced as to topics (e.g. symptoms vs governmental measures), but covers a wide range of domains.

METHODOLOGY

Multilevel Analysis

There is a proliferation of fallacy inventories associated with the various informal logic and rhetorical traditions (Hansen, 1996). This diversity has so far hampered systematic annotation of fallacies. Aristotle, for example, in his *Sophistical Refutations* (165b24-168a17), distinguishes fallacies dependent on the use of language and expression (*in dictione*), such as the fallacies of equivocation and ambiguity, from those not dependent on language (*extra dictione*), such as the fallacy of false cause. Pragmatic frameworks classify fallacies as infringements of the rules of an ideal critical discussion (van Eemeren & Grootendorst, 2004). Regardless of the chosen approach, the main issue at stake is the so-called *Fallacy Fork* (Boudry et al., 2015): cut-and-dry *compendia* of fallacies are unlikely to be found in real life discourse. To cope with this, we have adopted a bottom-up approach, with a focus on the analysis of the news articles in order to extract higher order insights. In this case, the expert annotator analyzed 40 fact-checked articles randomly picked from *Climatefeedback.org*, a platform that gathers a network of scientists engaged in sorting fact from fiction in climate change media coverage, and identified which fallacies have been called out through the comments of the reviewers. We intentionally focused on news related to a topic detached from COVID-19 but of public interest to check whether the resulting taxonomy is domain dependent or not.

As a starting point for our taxonomy of fallacies we adopted Tindale's (2007) framework, which gathers the most common fallacies discussed in the informal logic tradition. The resulting annotation schema includes 10 types of fallacies scattered into four main groups: fallacies related to the presence of (sufficient) arguments: *evading the burden of proof (EBP)*; fallacies pointing to the (un)intentional diversion of the attention from the issue at hand: *strawman (ST)*, *false authority (FAUT)*, *red herring (RH)*, and *cherry picking (CP)*; fallacies depending on the type of reasoning at play: *false analogy (FA)*, *hasty generalization (HG)*, *post hoc (PH)*, and *false cause (FC)*; fallacies related to the language used: *vagueness (VAG)*. The guidelines contain the description of the notion of fallacy and its relation to misinformation. Each fallacy is then defined, associated to an example, and accompanied by one or more critical questions, which have turned out to be useful means to evaluate arguments (Song et al., 2014). To offer a systematic and concise procedure, fallacies have been ordered starting from those having to do with the quantity of information provided, followed by those related to aspects external to the issue discussed; logical fallacies come into place after the other two classes are excluded. It is, in fact, not worth looking at the type of reasoning at play if the information conveyed in the arguments is not sufficient or irrelevant for the conclusion. The vagueness/ambiguity fallacy occupies the last position in the heuristics when all the other options have been considered. In this way, the annotator can go through the critical questions in a dyadic way, stopping when one of the critical questions is at stake:

Example (Kilpatrick & Fefferman, 2020)

Claim: "The WHO stated that asymptomatic spread of COVID-19 is 'very rare', therefore physical distancing and face masks are not necessary"

Fact-checker comment: "Imprecise: The scientific definition of the word "asymptomatic" refers only to a very small subset of infected people who never develop symptoms during the course of their infection. However, the public tends to interpret the word as also including pre-symptomatic individuals—those who are infected and not yet showing symptoms, but eventually go on to do so. The WHO official was not referring to pre-symptomatic individuals in her statement."

1. Does the news express an unassailable fact? Yes--->("REAL" NEWS); No--->
2. Are there any evidence/arguments apart from the author's personal guarantee? Yes--->3; No---> Evading the burden of proof
3. Is the reported evidence (if any) the only available? Yes--->5; No--->4
4. Is there any other data available which would bring to a different news? Yes--->5; No--->Cherry picking
5. Are the evidence/arguments relevant for the news? Yes--->6; No--->Red herring
6. Is the news criticizing/rebutting somebody else's opinion? Yes--->7; No--->8
7. Is the criticized/rebutted opinion misrepresented? Yes--->Straw man; No--->8

8. Does the news contain an appeal to authority (e.g. scientist, politician etc.)? Yes--->9; No--->10
9. Did the authority make the attributed claim? Yes--->10; No--->False authority
10. Is the authority a genuine and impartial source? Yes--->11; No--->False authority
11. Does the news contain the comparison between two different situations? Yes--->14; No--->13
12. Are the two situations alike for real? Yes--->13; No--->False analogy
13. Are the similarities/dissimilarities relevant to prove the truth of the news? Yes--->14; No--->False analogy
14. Is the news a generalization drawn from a sample? Yes--->15; No--->17
15. Is the sample representative of the population? Yes--->16; No--->Hasty generalization
16. Is the considered sample relevant to the circumstances of a present situation or does it constitute an exception? Yes--->”17; No--->Hasty generalization
17. Does the news express a causal relation (cause/effect) between situations? Yes--->18; No---> END (“REAL” NEWS)
18. Is it possible that the situations co-occur by coincidence? Yes--->Post hoc; No--->19
19. Could the situations be effect from separate or a common cause? Yes--->False cause; No--->20
20. Do concepts/words/phrases used in the news have multiple/vague/ambiguous meanings? Yes---> Language fallacy; No---> END (“REAL” NEWS)

Annotation of Types of Claims and Types of Sources

Semantic types of claims have been analyzed to identify features that make a standpoint persuasive or predict the types of arguments that are suitable to support them (Hidey et al., 2017). To investigate whether certain types of claims circulated through news are more or less likely to convey dis-misinformation and/or to be supported by fallacious arguments we have annotated the fact-checked news headlines using the following four main categories:

1. **Description (D):** The claim expresses a factual state of affairs, i.e., “there are x number of infections in London”; “the Oxford University lab has already produced a vaccine”.
2. **Prediction (P)**: The claim expresses a future state of affairs, i.e., “The economy will end up being destroyed”.
3. **Interpretation (I)**: The claim expresses an explanation of states of affairs, i.e., “The only reason why Italy has more cases, it is because they tested more”.
4. **Evaluation**: The claim expresses a more or less positive or negative judgement. Drawing from Liu (2012), evaluations are further classified as:
    a. **Evaluation-rational (ER)**: The claim expresses an opinion based on rational reasoning, non-subjective evidence or credible sources, i.e., “his phase 2 program is very solid”.
    b. **Evaluation-emotional (EE)**: The claim expresses an opinion based on emotional reasons and/or subjective beliefs, i.e., “I don’t’ like having to use a mask at all times”.

Our final layer of analysis consists in the annotation of the type of media source hosting the fact-checked news. Due to the inherent fluidity of the digital medium, taxonomies cannot rely merely on medium factors observed in computer mediated communication studies (Herring, 2007): with the rapid evolution of technological affordances features such as communication channels, synchronicity or message format are blurred. We have, thus, decided to draw upon social and situational factors in defining our types of media sources. More specifically, we have distinguished sources on the basis of the s*ocial practices*, “patterned ways of using technologies and shared knowledge systems” (Yates & Sumner, 1997) and *discourse communities*, groups of “reflexive actors with shared social practices and shared understandings of text types/genres, social contexts and communicative acts”, they give voice to: social media (e.g. Facebook, TikTok), broadcast media, blogs, scientific articles, governmental sources (e.g. Liverpool City Council website). Among broadcast media we have further distinguished broadcast media available through multi channels (e.g., Liverpool Echo), from those available digitally exclusively since they potentially reach out to different audiences.

Our multi-level analysis has been carried out by two undergraduate students with no previous background in argumentation theory or informal logic. They were introduced to fallacy theory and semantic types of claims as well as the task guidelines through a 90-minute training session. They were given the same set of news in CSV files and asked to identify:

a. type of semantic claim expressed in the headline;

b. type of source (e.g., social media); and

c. type of fallacies (if any) at stake.

The set of fact-checked news they assessed had been rated between completely "true" (signaling information) and completely "false" (signaling disinformation). They were also warned that a piece of news may contain more than one fallacy and asked to choose the one that is more clearly flagged by the fact checkers. Once the annotators completed the annotation process, we asked a rhetoric research specialist to go through the cases where the annotators disagreed and decide what label to retain (this produced the *golden standard annotation* set).

RESULTS OF THE ANALYSIS

Results of the Annotation (RQ1)

In order to evaluate the reliability of the annotations we have first calculated the inter-annotator agreement (IAA) using Cohen's kappa (κ) (Cohen, 1960) since we have two annotators. To interpret the kappa values, we have relied upon Landis and Koch's scale, obtaining the values in **Table 2**.

Table 2. Inter-annotator agreement metrics

Level of analysis	Kappa value	Type of agreement
Type of media source	0.68	Substantial
Type of semantic claim	0.43	Moderate
Type of fallacy	0.52	Moderate

The results show that while the types of media sources are easy to identify, the borders between types of claims and the types of fallacies are more blurred. This is not surprising since the *kappa* values are comparable with those obtained in tasks of similar complexity such as the annotation of argument schemes (Musi et al., 2016). It has to be remarked that our datasets constitute one of few annotated for fallacy type (Jin et al., 2022). Besides assessing the overall difficulty encountered by non-experts in using these analytical categories and offer a reliably annotated dataset, the main goal of the annotation was to understand what types of claims and what fallacies tend to be confused. On the one hand, different understandings of the semantics of news claims might trigger different decision-making processes: a piece of advice drawn from a claim perceived as descriptive is, for example, reasonably felt more reliable than one taken from a news expressing an interpretation. On the other hand, fallacies that are more challenging to identify are more likely to convey misinformation that is not recognized by the general public. To investigate these trends, we have built and analyzed the confusion matrices displayed in **Table 3** and **Table 4**.

Table 3. Confusion matrix for semantic type of claims

	D	P	I	EE	ER	Total
D	679	26	9	10	12	736
P	106	65	3	14	0	188
I	73	6	33	13	2	127
EE	5	1	0	40	0	46
ER	22	0	1	2	13	38
Total	885	98	46	79	27	1,135

Table 4. Confusion matrix for fallacy types

	EBP	ST	FAUT	RH	CP	FA	HG	PH	FC	VAG	NO	Total
EBP	73	2	4	3	2	0	1	0	0	1	13	99
ST	1	14	2	0	1	0	0	1	0	5	2	26
FAUT	2	1	11	2	2	0	2	1	0	2	5	28
RH	7	4	0	15	2	0	1	1	0	3	5	38
CP	8	2	3	3	39	0	3	1	1	2	16	78
FA	1	4	2	2	1	7	0	1	2	3	2	25
HG	15	5	4	1	10	1	47	1	2	6	7	99
PH	3	2	2	0	2	0	0	9	1	1	2	22
FC	0	0	0	1	1	0	1	0	2	1	0	6
VAG	7	1	3	5	11	1	6	1	0	51	13	99
NO	1	0	0	0	0	0	3	0	0	0	2	6
Total	118	35	31	32	71	9	64	16	8	75	67	526

The analysis of the confusion matrix in **Table 3** in comparison with the golden standard annotation revealed that the category "description" has been overgeneralized by one annotator, covering cases where the claim expressed instead a prediction, an interpretation, or an evaluation of the rational type. Zooming into those instances, it seems that the cases that have been confused present as recurrent features a modal verb (e.g., "fish tank additive may treat coronavirus", prediction confused with description) or reference to an authority (e.g., "Italy is hit hard, experts say, only because they have the oldest population in Europe"–interpretation confused with description; e.g. "the UK government no longer considers COVID-19 to be a high consequence infectious disease"–evaluation rational confused with description). This suggests that statements presented as possible states of affairs that could, thus, happen in the future, have the potential to be misinterpreted as factual at the moment of utterance; similarly, the ethos of authorities may lead to consider interpretations and evaluations as unassailable realities.

As far as fallacies are concerned, divergences between annotators are scattered across the full range, making it difficult to discern which pairs of fallacies tend to be confused more than others. However, it is clear that one of the two annotators had more difficulties in identifying the cherry picking fallacy. This is not surprising since the identification of arbitrary selection of sources requires a high degree of domain knowledge that is frequently hard to pinpoint, especially when available evidence has changed over time. It is, for instance, the case of the claim "Health authorities like the World Health Organization and the US Centers for Disease Control and Prevention discourage people from wearing face masks" which expresses, as pointed out by *Healthfeedback.org* an outdated as well as partial recommendation since "Health authorities initially discouraged the public from wearing face masks due to extreme shortages of surgical and N95 masks needed to protect healthcare workers. However, health authorities now recommend mask use by the public, as new evidence suggests that cloth face masks worn by the public effectively reduce COVID-19 transmission" (Teoh, 2020a).

The Misinformation Ecosystem (RQ2 & RQ3)

Solving cases of disagreement, the golden standard annotation has shed light on the misinformation ecosystem in our dataset of 1,135 news articles. As to the types of sources, social media represent the large majority (72%) and feature multi-modal content ranging from tweets to YouTube videos. This is in line with research showing the privileged role of social media as vehicles of fake news (Mahid et al., 2018). Broadcast media cover 19% of the news with a preference for multi-channels news (12%) available, for instance, on digital as well as paper versions of the New York Times. Blogs represent the 6% of the sources encompassing personal as well as group pages. Finally, governmental sources constitute 3% of the sources including both national and regional official venues.

Turning to the semantic type of claims, descriptions cover three quarter (68%) of the cases, either presenting conspiracy theories as factual (e.g., "the COVID-19 coronavirus disease is spreading quickly from gas pumps.") or advancing misleading information about a wide variety of topics (e.g., "Eating bananas is a preventative against the COVID-19 coronavirus disease"). Claims of the interpretative type (14%) tend to express in our dataset causal relations where negative state of affairs related to Covid-19 are presented as effects of other supposedly co-occurring state of affairs; the cause-effect relation is for the most directly marked through a causal connective or phrase (e.g., "the (COVID-19) cases are going up, but it's because the testing is going up."; "96.3% of the Italy's COVID-19 deaths were actually caused by other diseases"). Regardless the form of expression, this configuration confirms the need for humans to engage in abductive reasoning (i.e., most probable conclusion based incomplete information) when fronting uncertain scenarios, looking out for what they consider best possible explanations for situations otherwise difficult to understand. Predictions (9%), expressed with higher epistemic commitment, have mostly scope over future directions taken by the pandemic (e.g., "COVID-19 is here to stay" and "we need to accept that and be prepared to deal with COVID-19 long term") or outcomes of COVID-19 related policies (e.g., "the government in Oklahoma is planning to detain people unless they can show proof of vaccination"). Finally, among evaluative statements (9% overall), emotional evaluations (e.g., "while California is dying … Gavin (Newsom) is vacationing in Stevensville, MT!") outnumber (7%) rational ones (e.g., "we've tested more than every country combined") confirming that appeal to fearmongering is a common rhetorical strategy facilitating disinformation and misinformation spread.

When it comes to fallacies, the distribution across the 526 misinformation claims tagged as misinformation is visualized in **Figure 1**.

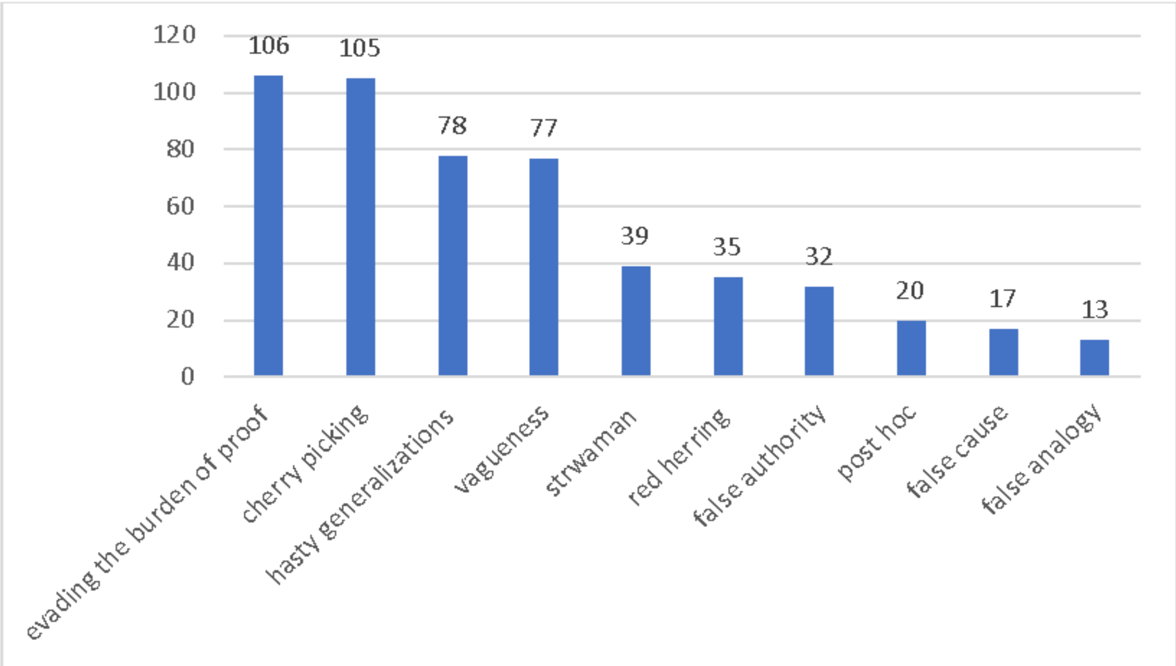

Figure 1. Distribution of fallacies in our dataset

The lack of sufficient arguments in support of a claim (*evading the burden of proof*) constitutes together with the *cherry picking* of evidence the most common fallacy in our sample, followed by generalizations drawn from a non-representative or balanced sample (*hasty generalization*) and the use of vague/polysemous language which allows for multiple interpretations (*vagueness*). Arguments which misrepresent a third party's opinion (*strawman*) or appeal to an inappropriate authority are also quite frequent together with arguments that are actually not relevant for the claim they support (*red herring*). Less common are the logical fallacies of post hoc, where a correlation is presented as a causation; *false cause*, where the wrong cause is attributed to an effect and *false analogy*, where a conclusion is drawn of the basis of similarities between two states of affairs which are not comparable.

It has to be noted that the total number of cases containing fallacies amounts to 522 instead of 526 as in the original annotation. This is because during the golden standard annotation process the expert annotator noticed that certain instances have been considered by the two annotators as instances of misinformation, while reporting no factual information to be classified as disinformation. Such cases stem from *Full Fact,* that does not include a fixed set of verdicts, and from cases labelled "Incorrect" in *Health Feedback* (instead of False as in other fact-checkers' truth barometers), as a further confirmation that the lack of a uniform set of verdict descriptors hinders the recognition of different types of fake news.

Even though the restricted size of our sample prevents us from drawing any correlation between the frequency of certain fallacy and the domain of the pandemic, it still suggests that the proposed taxonomy of fallacies bears descriptive power when it comes to the grey area of misinformation under COVID-19 since for each news rated as misinformation a fallacy has been identified by the annotators.

Analysis of Inter-Level Correlations (RQ4)

To investigate the backbones of the misinformation ecosystem, we analyzed the mutual distributions of our analytic categories throughout the dataset taking the golden standard annotation as a benchmark. Starting from the semantic level, we obtained a positive statistical correspondence ($\chi 2$ (36, n=514)=70.813, p=.0.000, with a medium effect size Cramer'sV=0.186) between the fallacy at stake and the type of claims that constitute the main headline of the news. Looking at the residuals and contributions with highest value (**Figure 2**), three main patterns stand out, namely interpretations * false cause; evaluation emotional * false analogy and prediction * evading the burden of proof.

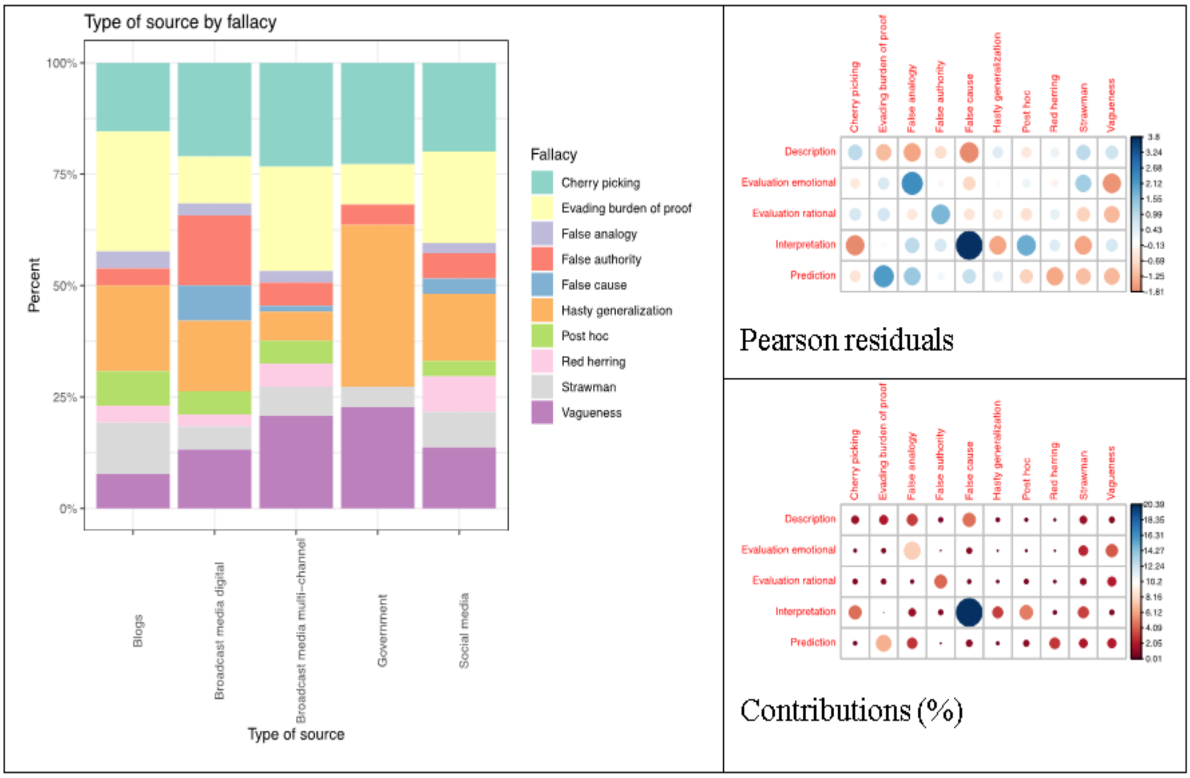

Figure 2. Types of claim per fallacy–proportions and Chi-squared residuals and contributions; percentages and residuals

While it is expected that flawed causal relations would be used as arguments for faulty interpretations (e.g., "There's a spike in [COVID-19] cases because there's a spike in testing") and that illegitimate comparisons would fire up evaluative statements with a subjective connotation, the association between predictions and *evading the burden of proof* is not intuitive. A predicament over a future state of affairs calls by default for evidence to be credible. Closer examination of these cases reveals that such predictions relate for the most to the decline of the virus with the warmer weather, drawing credence from people's hopes rather than facts.

We found no statistical correspondence between the type of claims and the type of source: $\chi 2$ (36, n=514)=22.544, p=0.127. Though it should be noted that descriptive claims dominate all sources and that government sources do not include evaluative claims (see **Figure 3**).

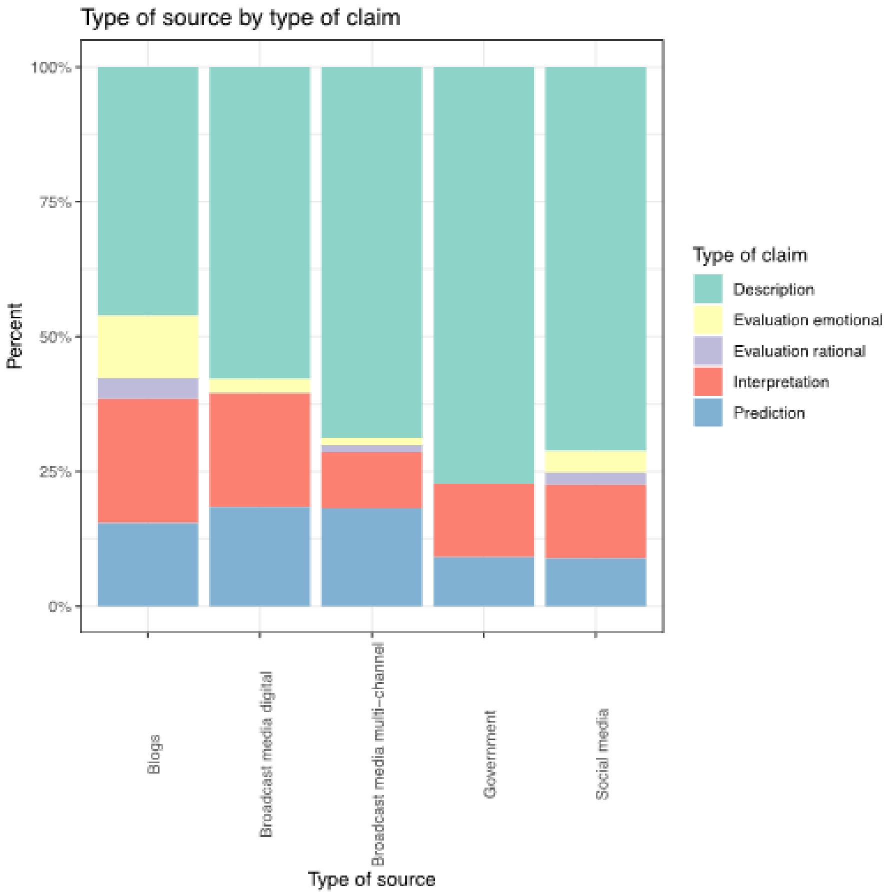

Figure 3. Types of claim per fallacy-percentages and residuals

Blogs appear to have more subjective types of claims (evaluation emotional and interpretations). Comparing just on subjective against non-subjective claims we find at statistically significant result at the $p<0.1$ level ($\chi2$ (4, n=514)=8.116, p=0.087, with a small effects size, Cramer's V=0.126). In this analysis (see **Figure 4**) blogs are the major contributor to the correspondence between factors. These results are limited by the nature of our sample. We speculate that blogs are more evaluative sources in line with their nature as digital spaces working as personal records. It is possible predictions tend to be preferred by broadcast media-multi-channels as the focus is on future impacts. Further analysis of a larger sample of cases will be needed to assess any consistent correspondence of sources and types of claim.

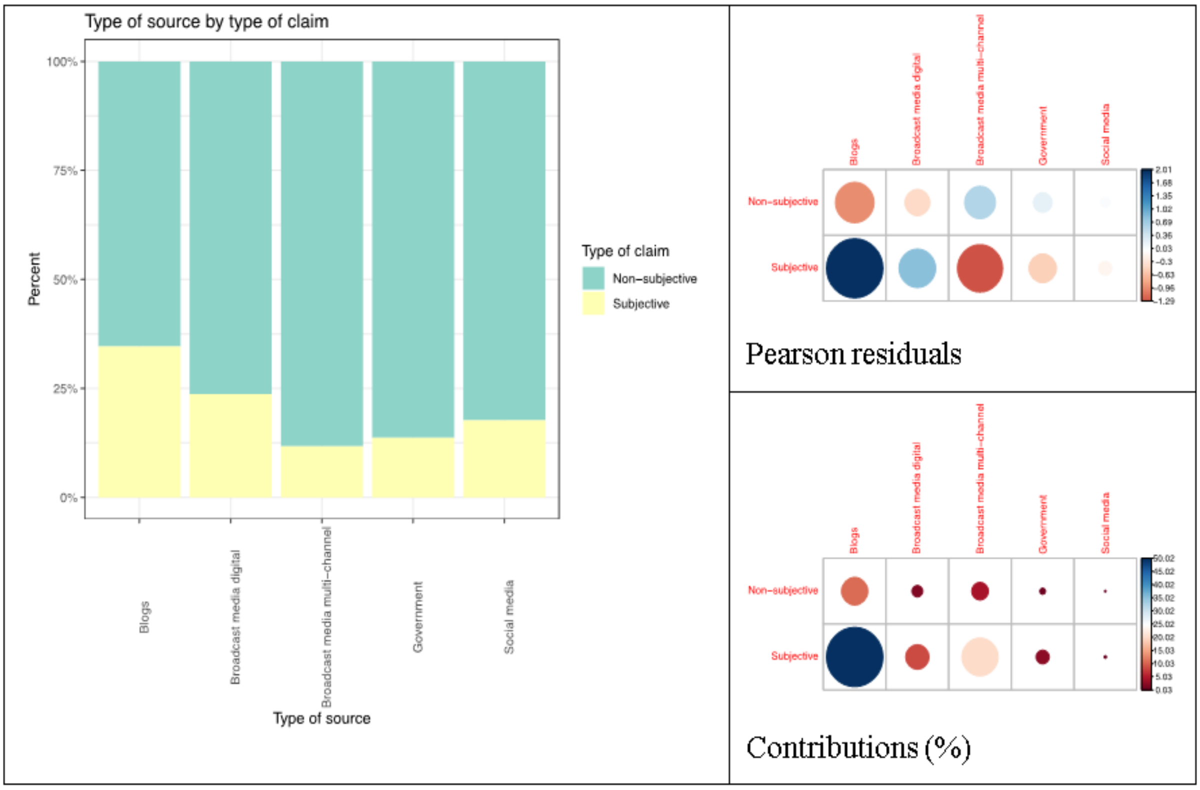

Figure 4. Types of claim per fallacy-percentages and residuals

The other variable that corresponds significantly with our classification of types of sources is the broad category of misinformation, disinformation and information ($\chi2$ (8, n=514)=33.139, p<0.000, small to medium effects size, Cramer's V=0.121). More specifically, while all the source types in our sample convey fake news as well as real news, social media and blogs constitute privileged channels for the spread of disinformation, while broadcast media and governmental official sources seem to be negatively correlated with blatantly false news. However, the trend is reversed when it comes to misinformation that bears positive residuals in correspondence with both broadcast media and government official sources (**Figure 5**).

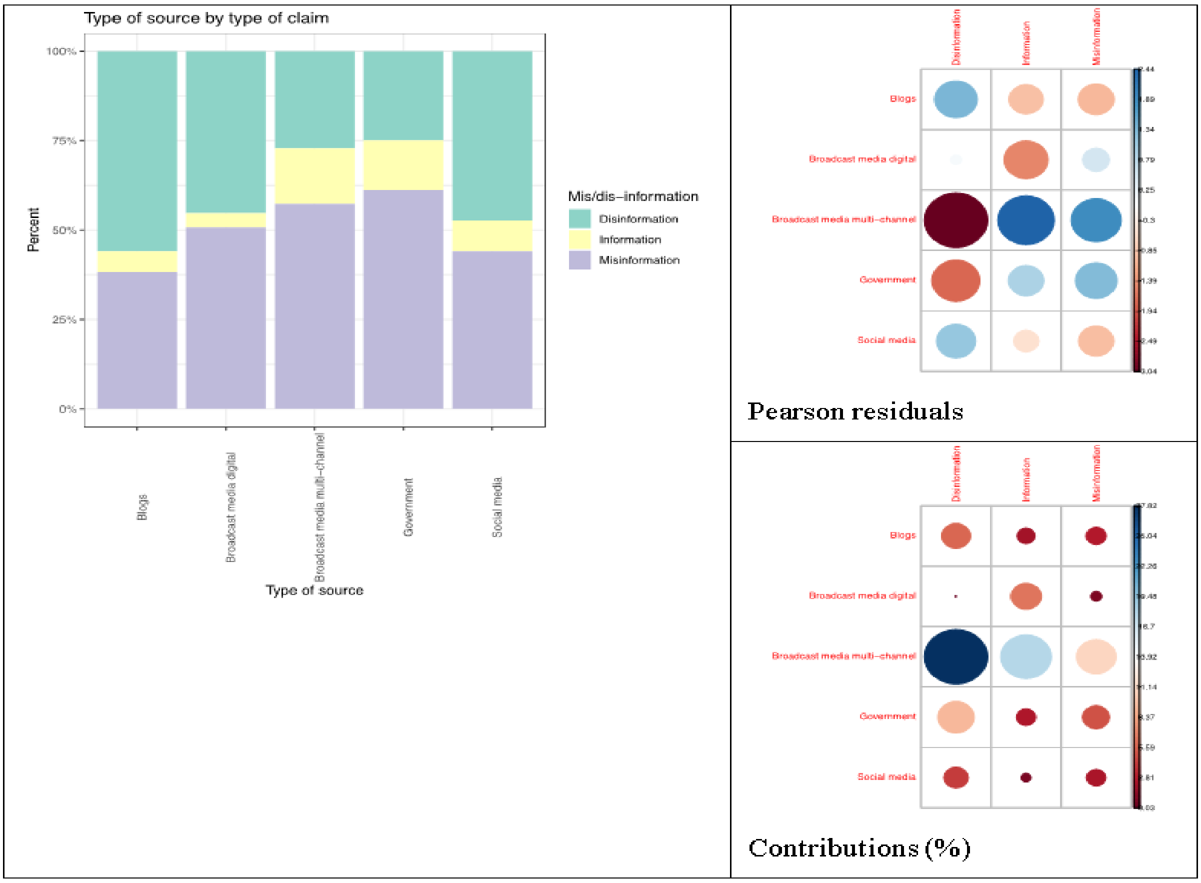

Figure 5. Types of source across the misinformation ecosystem-percentages and residuals

This trend partially aligns with results of studies showing that social media work as privileged vectors for the spread of conspiracy theories/completely false information (Allington et al., 2020; Li et al., 2020) and that governmental communications spread confusing information which might cause misinformation (Kyriakidou et al., 2020). The fact that misleading information can be spread by authoritative sources which are relegated to gatekeeping processes reveals a gap between intentions and outcomes in radically uncertain situations such as the pandemic. Looking at the distribution of fallacies (see **Figure 6**), we did not find any statistically significant correlation between fallacy classes and types of sources ($\chi2$ (12, n=514)=16.032, p=0.190). This result suggests that in crisis situations where epistemological differences between various publics (e.g. journalists, policy makers, citizens) happen to be conflated, the entire range of fallacious moves is potentially relevant across the board, regardless of the source. Official news media are in fact not exempt from the same type of fallacious arguments spread by social media and blogs. From a methodological perspective this trend also suggests that, even though qualitative categories such as that of fallacies allow us to operate a categorization of the misinformation behaviors across media sources, it is not possible to calculate "averages" and thus build reliable predictions without taking into account a variety of factors which go beyond single variables. From the qualitative analysis of our sample, it has, for example, emerged that a factor influencing the type of fallacious move at stake is the topic of the news: the *strawman* fallacy is mostly associated with news about policies rather than symptoms or cures for COVID-19. If a policy-related statement is a good candidate to become viral on social media or not, however, implies another set of factors which are hard to predict.

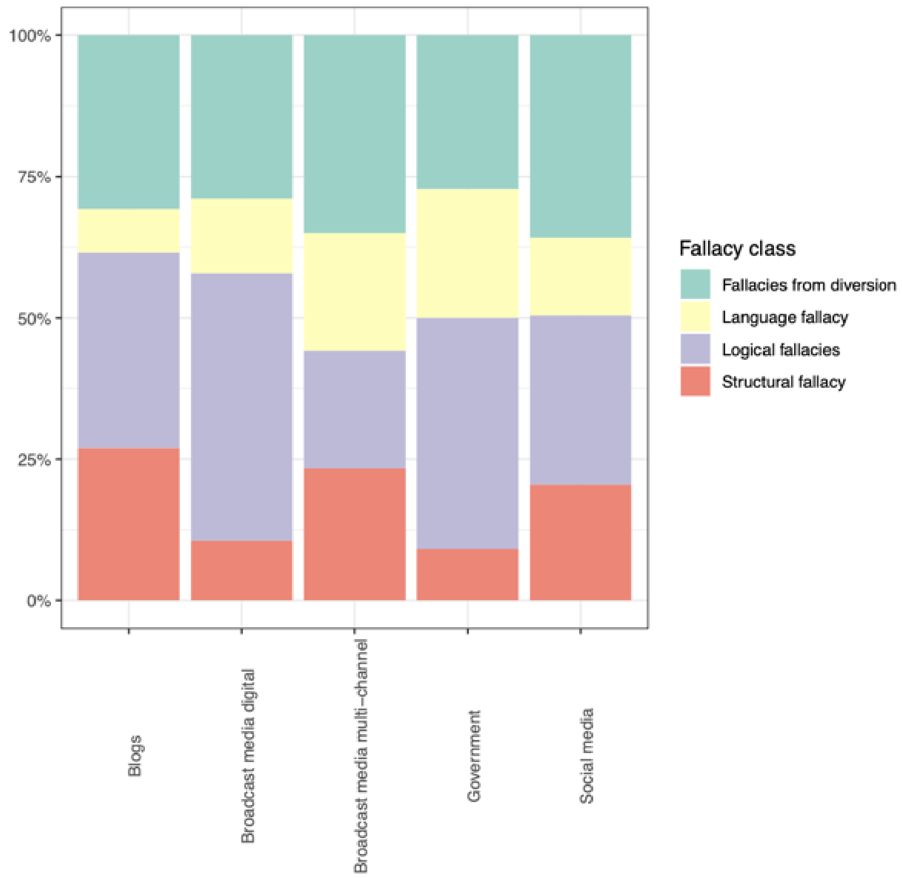

Figure 6. Fallacy class per type of source

CONCLUSIONS

In this study we address the phenomenon of fake news during the pandemic focusing on misinformation with the aim of contributing to its systematic identification. Fact-checking misinformation, that is, information which is misleading without necessarily containing false information communicated with the intention to deceive, imposes even more challenges than identifying disinformation. On the one side, automatic fact-checkers are currently unable to pick up information which may be factual, but misleading due to the lack of suitable training data; on the other hand, human fact-checkers struggle to keep up with the proliferation of information across digital media lacking a common truth barometer to flag the roots of misinformation. Drawing from the awareness that fact-checking is not always a matter of facts, but frequently a matter of how arguments supporting a news claim are built, we propose a discourse informed methodology to analyze misinformation leveraging critical thinking and, more specifically, fallacy theory.

The underlying theoretical starting point is that fallacies, defined as arguments that seem valid but are not, work as indicators of misinformation and provide more systematic explanations compared to mere labels as to why news might be misleading. To verify the explanatory potential of fallacies and investigate the COVID-19 misinformation ecosystem, we adopt a bottom-up approach through the corpus analysis of a dataset of 1,135 web scraped fact-checked news in English. A preliminary classification of the news according to the ratings shows that misinformation is more frequent than disinformation across the fact-checked dataset. We combine the annotation of fallacies, offering a novel heuristic procedure for their identification, with the annotation of type of sources and semantic type of news claims. While we obtain successful inter-annotator agreement metrics, the analysis of confusion matrices shows a tendency to overgeneralize the interpretation of news claims as descriptions even when a prediction, an interpretation or an evaluation is at stake, especially in the presence of a modal verb or a statement uttered by an authority (RQ1). Such results suggest that news headlines have to be more clearly framed to disentangle opinions from reported facts.

As to the fallacies, cherry picking seems to be the most difficult to identify and not surprisingly so since it requires a high level of epistemic vigilance and domain knowledge. The result of the golden standard annotation allows us to come up with a decalogue of fallacies which exhausts our misinformation dataset pointing to flows in the quantity and quality of arguments, the reasoning types at stake and the language used (RQ2). Besides working as indicators of misinformation that could be used as features to build systems for the automatic identification of misinformation, fallacies reveal the roots of misleading claims, being, thus, more informative than truth barometers proposed by current human fact-checking enterprises. In this way, understanding fallacies in social and broadcast media content may help people improve their digital literacy by learning how to cope with such online manipulations in the future.

The inter-level analysis between types of sources, claims and fallacies reveals that there are significant correlations between certain types of claims and fallacies as well as sources and that while social media are privileged sources for disinformation, misinformation is spread across the board, calling for more careful editorial processes in news production (RQ3 & RQ4). The attested patterns offer guidance to sharpen critical thinking when reading news, suggesting the need to keep epistemic vigilance high even when the sources are reliable news media outlets and to ask ourselves questions when reading the news pointing, for instance, to the presence of a sufficient number of arguments as well as the presence of correct inferences which do not, e.g., confuse correlations with causations.

Interestingly, different types of fallacies do not pattern significantly with different types of sources showing that crisis situations such as the pandemic where certainty is not an option constitute a challenging information environment for any kind of media. In such a post-truth scenario, audiences' digital literacy through critical thinking offers a very important response to counter the infodemic. We believe that our decalogue of fallacies constitutes a useful means to exercise audience's critical thinking towards reaching fake news immunity.

Author contributions: EM: main contributor to the design of the theoretical framework, the multi-level annotation, and the analysis of the results as PI on the research project; **MA:** contribution to the annotation process as golden standard annotator, coordinator of the annotators, and calculation of inter-annotator metrics; **EC:** contribution to theoretical framework (media literacy and digital literacy); **SY:** contribution to the statistical analysis of the results; and **KO:** contribution as a mentor in the analytic pipeline. All authors approve final version of the article.

Funding: This study is supported by UK Research and Innovation, Economic and Social Research Council, Grant No ES/V003909/1, Project "Being Alone Together: Developing Fake News Immunity", https://fakenewsimmunity.liverpool.ac.uk/

Declaration of interest: Authors declare no competing interest.

Data availability: Data generated or analyzed during this study are stored in the University of Liverpool Data Archive and will be made available from the authors on request.

REFERENCES

Abdulai, A. F., Tiffere, A. H., Adam, F., & Kabanunye, M. M. (2020). COVID-19 information-related digital literacy among online health consumers in a low-income country. *International Journal of Medical Informatics, 145*, 104322. https://doi.org/10.1016/j.ijmedinf.2020.104322

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2020). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine, 51*(10), 1763-1769. https://doi.org/10.1017/s0033291721000593

Amazeen, M. A., & Bucy, E. P. (2019). Conferring resistance to digital disinformation: The inoculating influence of procedural news knowledge. *Journal of Broadcasting & Electronic Media, 63*(3), 415-432. https://doi.org/10.1080/08838151.2019.1653101

Archila, P. A., Danies, G., Molina, J., Truscott de Mejía, A. M., & Restrepo, S. (2021). Towards Covid-19 literacy. *Science & Education*, *30*(4), 785-808. https://doi.org/10.1007/s11191-021-00222-1

Aufderheide, P. (1993). Media literacy. A report of the national leadership conference on media literacy. *Aspen Institute*. http://files.eric.ed.gov/fulltext/ED365294.pdf

Beaunoyer, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. *Computers in Human Behavior*, *111*, 106424. https://doi.org/10.1016/j.chb.2020.106424

Boudry, M., Paglieri, F., & Pigliucci, M. (2015). The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation*, *29*(4), 431-456. https://doi.org/10.1007/s10503-015-9359-1

Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*. http://www.primaonline.it/wp-content/uploads/2020/04/COVID-19_reuters.pdf

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review, 9*(2). https://doi.org/10.14763/2020.2.1481

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Fletcher, R., Kalogeropoulos, A., Simon, F. M., & Nielsen, R. K. (2020). Information inequality in the UK coronavirus communications crisis. *Reuters Institute*. https://reutersinstitute.politics.ox.ac.uk/information-inequality-uk-coronavirus-communications-crisis

Hamblin, C. L. (2022, March). *Fallacies*. Advanced Reasoning Forum.

Hansen, H. V. (1996). Aristotle, Whately, and the taxonomy of fallacies. In *Proceedings of the International Conference on Formal and Applied Practical Reasoning* (pp. 318-330). Springer. https://doi.org/10.1007/3-540-61313-7_82

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@ Internet*, *4,* 1.

Hidey, C., Musi, E., Hwang, A., Muresan, S., & McKeown, K. (2017, September). Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining* (pp. 11-21). https://doi.org/10.18653/v1/W17-5102

Hobbs, R. (2010). Digital and media literacy: A plan of action. A white paper on the digital and media literacy recommendations of the Knight Commission on the information needs of communities in a democracy. *Aspen Institute*. https://www.aspeninstitute.org/wp-content/uploads/2010/11/Digital_and_Media_Literacy.pdf

Jin, Z., Lalwani, A., Vaidhya, T., Shen, X., Ding, Y., Lyu, Z., Sachan, M., Mihalcea, R., & Schölkopf, B. (2022). Logical fallacy detection. *arXiv:2202.13758*. https://doi.org/10.48550/arXiv.2202.13758

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54*(1), 3-34. https://doi.org/10.3102/0002831216679817

Kasprak, A. (2020, December 10). No, mRNA COVID-19 vaccines do not 'alter your DNA' [fact-check of the news claim *mRNA vaccines are capable of altering or damaging human DNA*]. *Snopes*. https://www.snopes.com/fact-check/mrna-alter-dna/

Kilpatrick, A., & Fefferman, N. (2020, June 8). People who do not show COVID-19 symptoms can and do transmit it to others; physical distancing and face masks effectively reduce the risk of transmission [fact-check of claims on Facebook *the WHO stated that asymptomatic spread of COVID-19 is "very rare", therefore physical distancing and face masks are not necessary*]. *Healthfeedback.org*. https://healthfeedback.org/claimreview/people-who-do-not-show-covid-19-symptoms-can-and-do-transmit-it-to-others-physical-distancing-and-face-masks-effectively-reduce-the-risk-of-transmission/

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society*, *33*(2), 211-221. https://doi.org/10.1177/0163443710393382

Kyriakidou, M., Morani, M., Soo, N., & Cushion, S. (2020). Government and media misinformation about COVID-19 is confusing the public. *LSE COVID-19 Blog*. https://blogs.lse.ac.uk/covid19/2020/05/07/government-and-media-misinformation-about-covid-19-is-confusing-the-public/

Li, H. O. Y., Bailey, A., Huynh, D., & Chan, J. (2020). YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Global Health*, *5*(5), e002604. https://doi.org/10.1136/bmjgh-2020-002604

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Mahid, Z. I., Manickam, S., & Karuppayah, S. (2018, October). Fake news on social media: Brief review on detection techniques. In *Proceedings of the 2018 4th International Conference on Advances in Computing, Communication & Automation* (pp. 1-5). IEEE. https://doi.org/10.1109/icaccaf.2018.8776689

Musi, E., Ghosh, D., & Muresan, S. (2016, August). Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the 3rd Workshop on Argument Mining* (pp. 82-93). https://doi.org/10.18653/v1/W16-2810

Song, Y., Heilman, M., Klebanov, B. B., & Deane, P. (2014, June). Applying argumentation schemes for essay scoring. In *Proceedings of the 1st Workshop on Argumentation Mining* (pp. 69-78). https://doi.org/10.3115/v1/W14-2110

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital Journalism, 6*(2), 137-153. https://doi.org/10.17645/mac.v9i1.3331

Teoh, F. (2020a, August 4). Health authorities encourage the general public to wear face masks to reduce COVID-19 transmission [fact-check of claims on Facebook *health authorities like the World Health Organization and the US Centers for Disease Control and Prevention discourage people from wearing face masks*]. *Healthfeedback.org*. https://healthfeedback.org/claimreview/health-authorities-encourage-the-general-public-to-wear-face-masks-to-reduce-Covid-19-transmission/

Teoh, F. (2020b, December 19). The US national childhood vaccine injury act does not stop people from suing vaccine manufacturers [fact-check of the webpage https://archive.ph/OYp8S]. *Healthfeedback.org*. https://healthfeedback.org/claimreview/the-u-s-national-childhood-vaccine-injury-act-does-not-stop-people-from-suing-vaccine-manufacturers/

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3346–3359), Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806544

van Eemeren, F. H, & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616389

Vraga, E. K., Bode, L., & Tully, M. (2020). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research, 49*(2), 0093650219898094. https://doi.org/10.1177/0093650219898094

Yates, S. J., & Sumner, T. R. (1997, January). Digital genres and the new burden of fixity. In *Proceedings of the 13th Hawaii International Conference on System Sciences* (pp. 3-12). IEEE.

Yates, S., Carmi, E., Pawluczuk, A., Lockley, E., Wessels, B., & Gangneux, J. (2020). Me and my big data report 2020: Understanding citizens' data literacies: thinking, doing & participating with our data. *University of Liverpool*. https://www.liverpool.ac.uk/humanities-and-social-sciences/research/research-themes/centre-for-digital-humanities/projects/big-data/