

“Fake News” in Reformulated Messages

Towards Expanding the Toolset for Identifying Misinformation

Mitchell T. Welle, Marcin Koszowy

ABSTRACT In an age where information spreads faster than ever, the subtle manipulation of truth through rephrasing plays a pivotal role in amplifying misinformation. Starting from the observation that the spread of “fake news” may be significantly reinforced through reformulating a message for the sake of its *misrepresentation*, we seek to address the problem of the spread of “fake news” from the perspective of the rephrasing of news for purposes of misinformation. Given such a potentially dangerous role for *misuses of rephrasing*, the following research question arises: what is the relation between “fake news” and reformulated messages? This question will be addressed by analysing to what extent (i) definitions of “fake news” in the computer-science and philosophy-related literature, and (ii) recent linguistic studies of rephrase (as it is sometimes known), are helpful in identifying the main features of “fake news” as these relate to the latter. In this regard, we propose a research programme for addressing rephrase as a linguistic phenomenon—one that will serve as a tool for the study of communication in respect of “fake news.”

KEYWORDS fabrication; fake news; misrepresentation; mimicking; reformulating messages; rephrasing

Acknowledgements

We would like to acknowledge the fact that the work reported in this paper has been supported by the Polish National Science Centre under grant 2020/39/I/HS1/02861.

We extend our sincere thanks to the two anonymous reviewers for their constructive feedback and thoughtful engagement with our manuscript.

✉ Mitchell T. Welle, Laboratory of the New Ethos, Warsaw University of Technology, Poland

✉ Mitchell.welle@pw.edu.pl ☎ 0000-0003-2649-8831

✉ Marcin Koszowy, Laboratory of the New Ethos, Warsaw University of Technology, Poland

✉ marcin.koszowy@pw.edu.pl ☎ 0000-0001-5553-7428

Jill looked at the King: his mouth was open and his face was full of horror.
And then she understood the devilish cunning of the enemies' plan. By mixing
a little truth with it they had made their lie far stronger.

C.S. Lewis, *The Last Battle*

1. INTRODUCTION

In 2024, the President of the European Commission Ursula von der Leyen, at the World Economic Forum in Davos, Switzerland, stated: “For the global business community, the top concern for the next two years is not conflict or climate. It is disinformation and misinformation, followed closely by polarisation within our societies” (Von der Leyen 2024). So, in an age where information spreads faster than ever, the subtle manipulation of truth through *rephrasing* (or what linguists sometimes refer to as “rephrase”) plays a pivotal role in amplifying misinformation. This may give us pause to reflect and ask: “What is ‘rephrasing’?” The working definition of the latter runs as follows: “To say that a speaker rephrases is therefore to say that he or she means a contribution to be understood as connected to another contribution in a specific way” (Younis et al. 2023). This definition is built on contemporary research into rephrase (e.g. Visser et al. 2018; Konat et al. 2016; Koszowy et al. 2022). At the same time, the current state of research emphasizes the difficulties associated with the boundaries of the phenomenon: if rephrase is reformulation, where does it end? After all, two expressions of the same thought that differ completely in wording are difficult to classify as together constituting an instance of rephrase. Therefore, as has been stated, for example, by Younis et al (2023), an instance of rephrasing must be an operation on certain linguistic material that modifies it to a certain extent but not completely. However, this raises the following question: to what extent does this modification allow the output of this operation to be considered an instance of rephrase? In Koszowy et al. (2022), an attempt has been made to use semantic similarity as a measure, but counterexamples have been found showing that semantic convergence does not necessarily have to be related to reformulation. Therefore, in this article, we take as our starting point the definition of rephrase proposed by Younis et al. (2023). (The idea of rephrase will itself be elaborated in more detail in Section 4.)

Turning now to a timely example that uses a rephrased “fake news” headline to exemplify what we are talking about, we may consider the original news story “Netanyahu acknowledges Israel losing online “propaganda war,” should be doing more” from the *Times of Israel* (Freiberg 2025). The popular and (in)famous website *InfoWars* ran the rephrased headline “Netanyahu declares war on free speech as Israel’s propaganda

efforts falter” (2025). The news story on *InfoWars*, which itself is a story originally from the website “lifesitenews.com,” cites the *Times of Israel* as one of the many sources of the article. The *Times of Israel* article was the main article that lifesitenews.com drew its article from. The headline from the *Times of Israel* is relatively neutral. On the other hand, the *InfoWars* article includes phrases such as “falter” in place of “losing,” and employs the phrase “declares war” to intensify the propositional content in its rephrased title. In short, both titles are propositionally equivalent, but the rephrased title from *InfoWars* is linguistically intensified.

In this paper, we discuss the results of contemporary studies of rephrasing in argumentation and dialogue, treating these as a possible toolset for critically dealing with “fake news” and misinformation. Despite empirical evidence of the impact of repetitions on the spread of “fake news” (e.g. Hassan and Barber 2021; Diaz-Garcia et al. 2025), the intersection between communicative phenomena related to repeating or reformulating messages and the notion of “fake news” still remains underexplored. Here, we analyse the representative definitions of “fake news” in computer science and philosophy to reveal the key tendencies pertaining to the perception of “fake news” in these two research fields. Our analysis shows that the tendencies of reformulating or mimicking are embedded in those definitions. Given that the notions of reformulating and/or mimicking messages are implicated in techniques of rephrasing (in that rephrasing is a subcategory of mimicking), we hypothesize that the overlap between these two areas may provide a fruitful path for future research on the linguistic cues for identifying attempts at spreading “fake news” via rephrasing and reformulating techniques. Thus, we claim that recent models of rephrase in argumentation (Konat et al. 2016; Visser et al. 2018; Koszowy et al. 2022; Younis et al. 2023) can furnish a possible repertoire of tools for systematically studying and critically assessing rephrased “fake news.” This then leads on to an exploration of the key role of rephrasing “fake news” in the process of re-framing pieces of information for the sake of misinforming recipients.

Our initial observation is that the dissemination of “fake news” has become an integral challenge within the broader context of digital communication and media. This issue is particularly pertinent as such news leverages the pervasiveness of social platforms to manipulate public perception. One of the core tactics in “fake news” construction is the strategic use of rephrasing, where minor adjustments in language alter the truth value of statements while preserving their perceived factuality. This rephrasing technique subtly shifts public interpretation, often without causing overt contradictions, making it difficult to discern misinformation. These

observations prompt a deeper inquiry into the linguistic mechanisms that make such news persuasive.

Despite a growing corpus of research on “fake news” detection (e.g., Singhal et al. 2019; Tschitschek et al. 2017; Wang et al. 2018), there remains a critical gap in understanding how the rephrasing of statements contributes to the persuasive force of “fake news.” While previous studies (e.g., Hassan and Barber 2021) have concentrated on the broader identification of false information through algorithmic methods and empirical investigation focused on the effect of rephrasing in various domains (e.g., Koszowy et al. 2022; Younis et al. 2023), little attention has been paid to the cognitive and linguistic impacts of rephrasing in enhancing the believability of “fake news.” Specifically, the locutionary and illocutionary aspects of rephrasing as a rhetorical device in “fake news” production remain, to the best of our knowledge, underexplored. This lacuna highlights the need for an investigation that will bridge the gap between practical “fake news” detection on the one hand, and deeper philosophical considerations on the other.

Given this paucity of systemic research into reformulation strategies aimed at spreading “fake news,” our study seeks to address the following research question: what role does rephrasing play in the current structure of “fake news” in two representative research fields—namely, computer science and philosophy? Specifically, our research aims to explore how rephrasing techniques are *embedded* within representative definitions of both disciplines to distort or enhance the believability of “fake news.” By examining computational approaches to detecting rephrased misinformation and the philosophical implications of language manipulation, the study aims to offer an understanding of how rephrasing contributes to the structural integrity and rhetorical effectiveness of such news.

Our preliminary insights suggest that rephrasing plays distinct roles in the fields of computer science and philosophy, reflecting the differing focuses of these domains. In computer science, rephrasing is studied with a practical emphasis, as it typically has *real-world implications* for the detection and mitigation of “fake news” through algorithms and machine-learning techniques. These approaches aim to systematically identify “fake news,” thereby providing concrete tools for combating misinformation. On the other hand, philosophy engages with rephrasing in a *more speculative and theoretical* manner, focusing on the ethical and epistemological implications of how language can be manipulated to mislead or persuade. While both fields recognize the significance of rephrasing in the structure of “fake news,” computer science seeks actionable solutions, whereas philosophy probes the deeper implications of how and why rephrasing shapes our

understanding of epistemological and ethical concerns. More broadly, our goal here is rather to bring together two seemingly distant areas of research: one focused on rephrasing in pragmatics and argumentation theory, and the other dealing with the detection of “fake news” in the context of its increasing proliferation. The former is more philosophical and linguistic, while the latter, due to the computer systems and techniques used to detect such news, is definitely oriented towards computer science. Nevertheless, research on “fake news” detection procedures also has a clear philosophical component, because effective identification depends on an adequate definition of what “fake news” is (i.e. a fine-grained conceptual framework).¹

The goal of our research is to develop an understanding of how rephrasing contributes to the structure of “fake news.” To achieve this, we will study the definitions and conceptualizations of “fake news” in philosophy and computer science, specifically focusing on how rephrasing is treated within these frameworks. By analysing how computer science defines and detects rephrasing in such news through computational tools, alongside philosophical perspectives that explore the ethical and epistemological dimensions of rephrased misinformation, this research aims to bridge the gap by exploring the overlap between rephrasing and “fake news.”

For this purpose, we will be looking at notions of “fake news” in two research areas that significantly differ from each other, and so are complementary when it comes to obtaining an adequate conceptualization of such news from the point of view of the critical assessment of its role in communication, drawing out the components of rephrasing embedded in both sets of definitions. On the one hand, philosophy (especially epistemology and the philosophy of language, argumentation and communication) furnishes a theoretical framework for capturing the main features of “fake news.” On the other, computer science elaborates tools aimed at operationalizing the detection of (potential) “fake news” in its sphere of communication.

In this way, rephrasing becomes a form of deception that goes beyond the creation of falsehoods—it is about shaping these falsehoods to align with the expectations and cognitive biases of the audience. By crafting information that appears real, those behind “fake news” are able to amplify

1. To elaborate further, this is because these conceptualizations of “fake news” include, among other things, both an epistemological component (in that they address issues of truth and cognitive deception through “fake news”) and an ethical one (since they refer to the moral dimension of “fake news” at the level of the intentions behind its use and its effects). Hence, exploring the overlap between the study of rephrasing and research on “fake news” has directed us to philosophy and computer science as the representative disciplines typically and most visibly exhibiting this overlap.

the impact of their fabricated content, making it more persuasive and harder to discredit. This intentional rephrasing, coupled with the strategic presentation of information, plays a critical role in the effectiveness of “fake news,” ensuring that the falsehoods it contains are not only believable but also likely to be disseminated widely before they are debunked.

The overall idea behind the present paper is to *initiate moves* in the direction of laying a foundation for developing a comprehensive toolset for combating “fake news” from the perspective of computer science. The latter, via social media, offers a powerful and systematic method of identifying and neutralizing such news; however, some “fake news” articles manage to escape detection. We may deploy an analogy here: it is akin to a fishing net. Many of the obvious “fake news” articles are caught, but the smaller fish (i.e. the subtle “fake news” stories) manage to escape detection. One class of “fake news” articles that escape detection are rephrased news articles. The philosophical discussions surrounding “fake news” offer a more nuanced discussion of the nature of the latter, and we might find the discussion here fruitful for developing a more adequate definition of “fake news” for computer scientists. To give an example of how philosophy could potentially help computer science, we observe that many papers in the latter field define “fake news” as at the very least having the property of being intentionally false (see: Singhal et al. 2019; Desamsetti et al. 2023; Zhou and Zafarani 2018). However, as we can see from the *InfoWars* story referred to at the start, we can find ourselves dealing with an article that is not intentionally false, but where there is still an intuition that something of that type should somehow count as “fake news.” So why is there an intuition to the effect that the *InfoWars* example should count as an instance of such news?² A more nuanced account of this, taking into account aspects of rephrasing, may yet succeed in capturing that example within the net of “fake news” stories.

The overall structure of the paper, then, will be as follows: first, in Section 2, we explore key texts discussing “fake news” detection in computer science in relation to rephrasing; second, in Section 3, we examine the literature in philosophy with respect to its discussion of “fake news” as

2. It is worth mentioning that it is not the case *a priori* that *InfoWars* is “fake news.” We might treat it as a kind of institutional fact, to use Searle’s term. It just happens to be the case that as of this moment (2025), *InfoWars* is commonly taken to be the paradigmatic example of this (see Dentith 2016). Does this always have to be the case? No. It certainly is the case that in some possible world Alex Jones, the head of *InfoWars*, speaks as though he were trained to deliver the news on NPR and adheres to the news standards of the Associated Press (AP) before publishing news articles. Additionally, is everything that *InfoWars* publishes or reports on false? No. But should we be sceptical about stories published on *InfoWars*? Yes.

it relates to rephrasing; third, in Section 4, by inspecting the similarities and differences in these fields we seek to bridge the gap between them, surveying the overlap between rephrase and “fake news” in order to propose specific areas of philosophical and linguistic research into rephrasing; lastly, in Section 5, we conclude by envisaging possible future work. These steps will, we hope, enable us to offer a new perspective on rephrasing and “fake news” studies, legitimizing the category of rephrasing as one of the key concepts needed for capturing the linguistic manifestations of “fake news.” In this way, our research programme aims to capture both theoretical dimensions of rephrasing as they relate to this phenomenon.

2 DEFINITIONS OF “FAKE NEWS” IN COMPUTER SCIENCE

In exploring the role that rephrasing plays in definitions of “fake news,” we have limited our investigation to the fields of philosophy and computer science. This decision is driven by their complementary strengths in addressing the nuanced ways that rephrasing can be used to manipulate information. Philosophy provides a critical theoretical framework for understanding how language, including rephrasing, can be used to obscure truth or distort reality. It allows us to examine the ethical implications of “fake news” as a tool for deception, exploring the philosophical foundations of truth and the societal impacts of manipulated information.

In seeking to gather together the necessary computer-science definitions related to “fake news,” our approach has been designed to ensure relevance during the selection process. First, we identified a series of relevant, and commonly cited and recent, peer-reviewed articles falling within the field of computer science that explicitly addressed “fake news” or focused on methodologies for “fake news” detection. These articles served as the primary sources for extracting the definitions of “fake news.” We screened peer-reviewed work between 2015–2025 in venues indexed by ACM Digital Library, IEEE Xplore and major journals/conferences; for philosophy, we screened journal articles and chapters in epistemology, as well as the philosophy of language, argumentation and communication. Second, from each article, we isolated and recorded the specific definitions of “fake news” utilized by the authors, ensuring that these were contextually aligned with the overall research objectives of the study. Third, we compiled these definitions into a table, providing a comparative framework for analysis. Lastly, we examined the definitions for recurring features or thematic patterns, which were then highlighted and categorized within the table to illustrate commonalities and distinctions across the selected literature. We drew from these features or thematic patterns elements aimed at identifying linguistic

manifestations of certain instances of rephrasing that may have the effect of “fake news.” This approach allowed for a comparative understanding of how such news is conceptualized within the domain of computer science.

Table 1: Computer science definitions of “fake news”: major tendencies

#	Author(s)	Falsehoods	Fabrication	Intent And Purpose	Ways of mis-leading people	Verification and evidence	Harm and other normative factors	Undefined
1.	Dong et al. (2023)							X
2.	Qian et al. (2018)							X
3.	Ruchansky et al. (2017)	X	X					
4.	Tschiatschek et al. (2017)	X			X			
5.	Alam et al. (2022)	X		X	X		X	
6.	Kou et al. (2022)	X			X			
7.	Nabov et al. (2021)				X		X	
8.	Zhou et al. (2020)	X		X				
9.	Ajao et al. (2019)	X		X	X	X		
10.	Khattar er al. (2019)	X	X			X		
11.	Singhal et al. (2019)	X		X		X		
12.	Wang et al. (2018)	X	X	X		X		
13.	Conroy et al. (2015)	X		X				
14.	Farajtabar et al. (2017)	X		X			X	
15.	Berrondo-Otermin & Sarasa-Cabezuelo (2023)	X		X	X	X	X	
16.	Wu & Rao (2020)	X			X			
17.	Desamsetti et al. (2023)	X		X			X	
18.	Fifita et al. (2023)	X		X				
19.	Shu et al. (2017)	X		X				
20.	Zhou & Zafarani (2018)	X		X				
21.	Zhang & Ghorbani (2020)	X		X	X		X	
22.	Jain & Kasbe (2018)	X			X			
23.	Perez-Rosas et al. (2017)	X	X	X		X		
24.	Oshikawa et al. (2018)	X			X			
25.	Parikh & Atrey (2018)	X		X				
26.	Sharma & Singh (2024)	X		X	X		X	

Table 1 presents a detailed overview of the major tendencies found in definitions of “fake news” within the computer-science literature. To compile this table, we analysed 26 representative articles, each of which directly addresses either “fake news” itself or its detection. The definitions are categorized into six key themes: Falsehood, Fabrication, Intent and Purpose, Ways of Misleading People, Verification and Evidence, and Undefined. These themes were inductively drawn from the definitions of “fake news,” prior to our identifying linguistic manifestations of certain instances of rephrasing that may have the effect of “fake news.” Falsehood and Fabrication emerged as fundamental components, frequently mentioned across the literature, underscoring the fact that “fake news” is often built upon deliberately false or fabricated information. The category of Intent and Purpose highlights the intentional deception that characterizes “fake news,” distinguishing it from accidental misinformation. The instances of Ways of Misleading People identified by us, including disinformation and satire, emphasize the diverse strategies employed to deceive audiences. Additionally, the role of different media platforms, such as social media, in the spread of “fake news,” is a prominent theme, as is the challenge of verifying the authenticity of information presented as news. This analysis reveals not only the consistency in how “fake news” is defined within the field of computer science, but also the emphasis on intentionality and the mechanisms of dissemination.

The structure of our discussion of these categories will be as follows: we will first describe the data obtained by giving the relevant statistical information, and then proceed to outline the elements of rephrasing identified in the features of various definitions of “fake news.” The goal is to see how rephrasing manifests itself within these features.

The first category is Falsehood. Falsehoods are a fundamental aspect of the definitions of “fake news” within the computer-science literature. This category, which includes terms such as “false,” “untrue,” and “falsified,” was identified in 23 out of 26 articles analysed, representing 88.4% of the total. This near-universal acknowledgement—as might well have been expected, given that the term suggests some form of falsity in one way or another—underscores the central role of falsehood in the conceptualization of “fake news.”

When we turn to the act of rephrasing, applied to false information, this serves as a powerful tool in the construction of “fake news.” Building on previous research into rephrasing (Konat et al. 2016; Visser et al. 2018; Koszowy et al. 2022; Younis et al. 2023), we may say that rephrasing in the context of “fake news” is not merely a simple alteration of words but

a deliberate effort to replicate the tone, style and structure of legitimate news sources (Dentith 2016). By doing so, creators of “fake news” aim to enhance the believability of their content, making it more difficult for the audience to distinguish between authentic and false information. The subtle manipulation of language allows the fabricated content to blend seamlessly with genuine news, exploiting the trust that readers place in familiar journalistic formats. This mimicry extends beyond mere language to include the visual and contextual elements of news, such as headlines, imagery, and even the use of credible-sounding sources, all designed to reinforce the illusion of authenticity.

The category of Fabrication stands out as notable within the computer-science literature. It included the term “fabrication” or “fabricate,” and was identified in 5 of the 26 articles (representing 19.2% of the total). The concept of fabrication is interesting because it shows an overlap between falsehood and the intentional aspect of “fake news,” distinguishing it from other forms of misinformation, such as rumours or inadvertent errors, as well as stories that are simply false (e.g. Ruchansky et al. 2017; Khattar et al. 2019; Farajtabar et al. 2017; Wang et al. 2018; Perez-Rosas et al. 2017). The act of fabricating involves creating false information intentionally, often with the aim of deceiving or misleading the audience. This deliberate intent is crucial for differentiating “fake news” from mere mistakes or misunderstandings. This distinction allows us to distinguish between obvious examples of “fake news” and mere mistakes. For example, many have the intuition that the highly cited *InfoWars* story about how the 2012 Sandy Hook school shooting in the US was staged is a paradigmatic example of “fake news” in the digital era (see Collins and Eaton-Robb 2022)—as distinct from being a simple case of confusion over some relatively subtle philosophical terminology employed in the Dennett obituary published by the *New York Times* (see Kandell 2024).

Fabrication, in the context of “fake news,” frequently involves rephrasing as a method for constructing misleading narratives. Rephrasing plays a crucial role in how fabricated content is framed and presented to the public. By subtly altering legitimate information or reconstructing false statements in a manner that mimics the tone and structure of credible sources, fabricators can create content that appears authentic, despite its deceptive intent. The process of rephrasing fabricated information often involves changing emphasis, removing critical context, or introducing emotionally charged language, all while maintaining an outward appearance of legitimacy. This strategic manipulation of language allows fabricators to enhance the believability of their falsehoods, making them more difficult

for audiences to detect and scrutinize. Thus, rephrasing acts as a vehicle for delivering fabricated content in a way that maximises its persuasive impact while concealing its intentional falsity. This connection between fabrication and rephrasing underscores the importance of developing tools that can identify subtle linguistic cues to better detect and counteract “fake news.”

Intent and Purpose captures the deliberate motivations behind the dissemination of false information. This category, which includes terms such as “intention to deceive,” “deliberate,” “purposeful” and “intentional,” was identified in 14 out of the 26 articles analysed, representing 61.5% of the total. This significant proportion underscores the importance of understanding the underlying motivations behind “fake news” in both academic and practical contexts. Unlike accidental misinformation, “fake news” is characterized by a conscious effort to mislead or deceive the audience. For instance, Alam et al. (2022), Singhal et al. (2019), Conroy et al. (2015), Fifita et al. (2023) and Parikh and Atrey (2018) highlight the purposeful nature of “fake news.” This intentionality is a key factor that differentiates “fake news” from other types of misleading information, such as rumours or unintentional errors.

To show that we can infer an intent to generate “fake news,” we can look at how rephrasing operates as a linguistic mechanism to subtly reshape stories to influence public opinion, promote a political agenda, or generate financial gain. This rephrasing often involves subtle changes in wording, tone, or emphasis, that significantly alter the perception of the information without making it immediately obvious that it has been manipulated. For instance, the title from the unconventional news website *Zero-Hedge*, “US Women’s Life Expectancy is the Lowest among Developed Nations” (Durden 2024), appears alarming, but upon closer analysis is misleading. The article provides a list of 10 developed countries, placing the United States at the bottom (at 80 years). However, two critical issues arise: first, the list excludes Slovakia (see WHO 2024), a country classified as developed (see UN 2014), where women’s life expectancy is 78 years. Secondly, the original article cited by *Zero-Hedge* refers to “high-income countries,” a term with a specific technical meaning, while the rephrased title substitutes it with “developed countries,” a term that carries broader and more urgent implications. This strategic rephrasing shows the intention of creating more urgency than the original article suggested, while still appearing to rely on solid evidence (this is also the case in the *InfoWars* example in Section 1. This example highlights how rephrasing operates as a linguistic mechanism to subtly reshape narratives, exploiting the trust in seemingly factual claims, while we nevertheless can infer an underlying intent behind the rephrased content.

Ways of Misleading People focuses on identifying the various forms that “fake news” can take. This category, which includes terms such as “misinformation,” “disinformation,” “hoaxes,” and “rumours,” was identified in 9 out of the 26 articles analysed, accounting for 42.3% of the total. The inclusion of terms like “misinformation” and “disinformation” in the definitions emphasizes the spectrum of accuracy and intent within “fake news.” Tschatschek et al. (2017) and Alam et al. (2022) discuss these distinctions, noting that misinformation typically involves the unintentional spread of false information, while disinformation is deliberately misleading. This differentiation is crucial for understanding the motivations behind the dissemination of “fake news,” as disinformation often carries malicious intent, while misinformation may arise from ignorance or error. Additionally, the category encompasses various formats of “fake news,” such as satire and hoaxes. For instance, Ajao et al. (2019) include satire as a form of “fake news,” recognising that while it may use humour and exaggeration, it can still mislead audiences who take the content at face value. This highlights the complexity of “fake news,” where even content meant as a joke can contribute to misinformation if not clearly identified as satire.

The Misinformation and Types of News category within “fake news” highlights the various forms it can take, and underscores how rephrasing plays a crucial role in shaping these forms to deceive or mislead. The inclusion of terms like “misinformation” and “disinformation” in the definitions emphasizes that “fake news” exists on a spectrum of accuracy and intent, where rephrasing can shift the meaning and perception of the information presented. Rephrasing allows creators to manipulate content in a way that aligns with their specific intent—whether to mislead unintentionally, as in the case of misinformation, or with malice, as with disinformation.

An effective analogy to describe this problem is the popular child’s game commonly known as “the Telephone Game,” where a message is whispered from one person to another in a line. By the time the message reaches the last person, it is often significantly distorted, bearing little resemblance to the original. In the context of “fake news,” disinformation acts similarly to the Telephone Game but with intentional distortion: the original information is deliberately rephrased at each stage to create a narrative that suits a particular agenda. When we turn to the news, we can see that the use of the phrase “swift-boating” became popular during the 2004 US presidential campaign involving John Kerry and George W. Bush. The phrase was a half-truth repeated to great effect and, arguably, caused Kerry to lose the

election (see Major and Andersen 2016).³ This kind of manipulation can introduce subtle inaccuracies or misleading interpretations, transforming what might have been factual content into something that is false yet appears credible.

Misinformation, though typically spread without harmful intent, can also resemble the unintended distortions in the Telephone Game. As information is passed along without proper verification, rephrasing may simplify, exaggerate, or alter details, leading to a message that is different from the original, even if the distortion is unintentional. This analogy highlights how rephrasing, even without malicious intent, can still lead to significant misunderstandings and the spread of inaccurate information. Additionally, the category includes various formats of “fake news,” such as satire and hoaxes, where rephrasing also plays a key role. Satire, for instance, relies heavily on exaggeration and humour, but when its satirical nature is not made clear, rephrasing can blur the line between joke and reality, leading some audiences to take the content at face value. Similarly, hoaxes often involve rephrasing factual information in a way that creates a completely false narrative, crafted to deceive the audience for the sake of entertainment, financial gain, or other motives.

For each kind of variation of “fake news,” it becomes apparent that rephrasing plays a significant role as a tool in its production. Whether the intent is to deceive, entertain, or simply provoke a reaction, rephrasing serves as a tool to reshape information into various types of “fake news,” each with its own impact on the audience. Understanding how rephrasing contributes to the complexity of “fake news,” much like the distortions in the Telephone Game, helps to illuminate the challenges in identifying and countering it in all its forms.

Turning to the Verification and Evidence category, this distinguishes “fake news” from other types of information, emphasising the presence or absence of supporting evidence and the verifiability of the content. This category, which includes terms such as “lack of evidence,” “unverifiable” and “false evidence,” was identified in 3 out of the 26 articles analysed, representing 26.9% of the total amount. The emphasis is on the unverifiable

3. For some more context: The term comes from Swift Boat Veterans for Truth (later Swift Vets and POWs for Truth), a “527” group that, in August 2004, ran TV ads and promoted the book *Unfit for Command* attacking Democratic nominee John Kerry’s Vietnam War record. (Kerry had commanded a Navy “swift boat.”) Many of the allegations were contradicted by Navy records and eyewitnesses and widely judged to be unsubstantiated. However, the ethotic attack achieved its greater strategic purpose regardless of the actual truth value of the main claim of the book and TV ads.

nature of “fake news.” For instance, Zhou et al. (2020) and Perez-Rosas et al. (2017) highlight the fact that a key characteristic of “fake news” is its lack of verifiable evidence. This absence of reliable sources or a factual basis makes it challenging to authenticate any information that, typically, would rely on verifiable data and credible sources. Khattar et al. (2019), interestingly, examines the concept of false evidence, where “fake news” lacks verification and may present fabricated or misleading evidence to support its claims. Here, it might be useful to draw a distinction to the effect that all false evidence is fabricated or misleading, but not all fabricated or misleading claims count as false evidence. It is worth noting that one issue with this category is that some legitimate news sources are difficult or impossible to verify. One might consider, for example, a news article that uses whistle-blower information or depends on confidential sources.

Rephrasing plays a pivotal role in how “fake news” navigates the Verification and Evidence category, particularly by obscuring the lack of verifiable information, or fabricating evidence to support false claims. When creators of “fake news” manipulate content, they often engage in rephrasing to make unverifiable information appear more credible in ways difficult to detect. This manipulation underscores the implicit imperative to cultivate robust critical thinking skills and dispositions amongst news consumers.

The Harm and Normative Factors category focuses on the ethical and societal implications of “fake news,” highlighting the potential damage caused by the dissemination of false information. This category includes terms such as “harm,” “unethical,” “normative judgments” and “negative impact.” It was identified in 5 out of the 26 articles analysed, representing 26.9% of the total amount. This number is noteworthy in the field of computer science, as it seems desirable for this normative dimension to be captured in the concept of “fake news.”

Harm is a central theme in this category, reflecting the various ways “fake news” can negatively impact individuals and society. For instance, Ruchansky et al. (2017) and Sharma and Singh (2024) discuss how “fake news” can lead to real-world consequences, such as social unrest, public panic, and even violence. These studies highlight the potential for “fake news” to exacerbate conflicts, spread fear, and undermine social cohesion, showcasing the broader societal harm that can result from the spread of misinformation. Alam et al. (2022) addresses the ethical dimensions of “fake news,” describing it as inherently reflecting an intention to deceive and manipulate the public. This ethical violation is particularly concerning when “fake news” is used to influence public opinion or manipulate political outcomes, as it undermines democratic processes and erodes public trust

in media and institutions. The deliberate spread of false information with the intent to mislead is widely regarded as an unethical practice, raising important questions about the responsibilities of information providers and the need for accountability.

The use of “normative judgments” and the discussion of “negative impact” further highlight the moral considerations associated with “fake news.” Berrondo-Otermin and Sarasa-Cabezuelo (2023) emphasize the normative implications of spreading misinformation, noting that “fake news” often involves making normative claims that are intended to shape public attitudes or behaviours. Additionally, Desamsetti et al. (2023) discusses the broader negative impacts of such news on public discourse, pointing out that the spread of false information can lead to a general degradation of the quality of information available to the public. This degradation can have long-term consequences, including the erosion of informed public debate and the weakening of societal resilience against misinformation. The authors highlight the need for stronger measures to combat “fake news” and protect the integrity of public discourse.

Rephrasing plays a critically important role in the Harm and Normative Factors category, particularly in respect of how “fake news” is crafted to inflict ethical and societal harm. The act of rephrasing in such news causes harm by making “fake news” nearly indistinguishable from real news. Also, the manipulation of language through rephrasing can amplify the negative impact of “fake news” by making harmful content appear more credible or persuasive. For instance, the ethical violation inherent in “fake news” is often deepened by the way information is rephrased to deceive or manipulate the public. Rephrasing can involve twisting facts, selectively omitting context, or framing information in a way that heightens emotional responses, all of which contribute to the unethical nature of “fake news.”

The term “harm” is central to this category, and rephrasing is frequently used to exacerbate the potential damage caused by “fake news.” For example, rephrased content may be designed to incite fear or anger, leading to real-world consequences such as social unrest or violence. By carefully choosing words and altering the presentation of information, creators of “fake news” can intensify the harmful effects on individuals and society, making the false information more likely to provoke extreme reactions. This manipulation of language not only spreads misinformation, but also actively contributes to the deterioration of social cohesion and public trust.

Rephrasing, here, also plays a role in the normative judgments associated with “fake news.” When such news is rephrased to make normative claims—statements that express values or prescribe actions—it can shape

public attitudes and behaviours in ways that align with the deceptive intent of its creators. This rephrasing can lead to the spread of harmful ideologies or the manipulation of public opinion, particularly when the language used is crafted to resonate with existing biases or fears. The ethical implications of such rephrasing are significant, as it involves deliberately steering public discourse in a direction that undermines the integrity of information and democratic processes.

Moreover, the broader negative impacts of “fake news” on public discourse, as discussed by authors such as Desamsetti et al. (2023), are often magnified through the strategic rephrasing of content. By presenting misleading information in a way that appears authoritative or aligns with popular narratives, rephrased “fake news” can degrade the overall quality of public debate. This degradation erodes societal resilience against misinformation, as the public becomes less able to distinguish between truth and falsehood in a landscape where rephrased “fake news” is prevalent.

The last category is the Undefined one. Some papers left the concept of “fake news” undefined (Dong et al. 2023; Qian et al. 2018). This absence of any definition is itself notable, because it assumes that the definition of “fake news” is so well understood that there is no need to supply one.

To conclude, our investigation into definitions of “fake news” in computer science has underscored the centrality of rephrasing as a mechanism for subtle misinformation. Central to this phenomenon is the use of falsehoods and fabrication, where strategic linguistic adjustments create an illusion of authenticity while embedding misinformation. A key tactic involves mimicking credible sources, replicating the tone, style and structure of legitimate journalism to exploit audience trust. Rephrased content often reflects deliberate intent and purpose, aligning language with specific objectives such as promoting propaganda or influencing public opinion. These manipulations frequently involve selective omission, emphasis, and pragmatic shifts that subtly alter meaning without overt contradiction. Furthermore, rephrasing serves to obscure verification and evidence, either by masking unverifiable claims or fabricating support to enhance believability. The societal and ethical impact of such rephrasing is significant, as it amplifies harm by provoking emotional reactions and exacerbating divisions. Detection tools in computer science focus on identifying linguistic patterns, structural mimicry, and shifts in tone to counteract the persuasive force of rephrased misinformation. Together, these elements underscore the nuanced and deceptive nature of rephrasing as a tool in fake news.

3. DEFINITIONS OF “FAKE NEWS” IN PHILOSOPHY

In this section, we examine the philosophical discourse surrounding definitions of “fake news,” focusing on how various philosophers have sought to conceptualize this phenomenon. While much of the existing work on “fake news” has centred around its detection and mitigation through technical means (as in computer science), philosophy offers a more nuanced and theoretical exploration of its ethical, epistemological and linguistic dimensions.

“Fake news” raises deeply philosophical issues. The topic raises significant questions about truth, epistemology, ethics, and the impact of information on society in the Digital Age. Below, we will explore the philosophical dimensions of “fake news,” focusing on how various theoretical frameworks can help us understand and address this complex problem as it simultaneously relates to computer science. Philosophers have long been concerned with the nature of truth and the ethical implications of communication, making the study of “fake news” particularly relevant to contemporary philosophical discourse, which in turn might provide useful contemporary tools for identifying such news. However, it is worth noting that there is no single agreed-upon definition of “fake news” (albeit that such a state of affairs is hardly an uncommon occurrence in philosophy). In fact, there are some philosophers (e.g. Habgood-Coote 2019; Musi and Reed 2022) who have either asserted that defining “fake news” is a fool’s errand, or sought to side-step it entirely. Nevertheless, the emphasis here is not so much on technological tools being used to identify “fake news,” as in computer science. Rather, what we encounter here is a deeper discussion about the nature of falsity, with more attention paid to such subtle aspects of “fake news” as the copying or mimicking of legitimate news.

Now we will turn to the analysis of selected definitions of “fake news” in the philosophical literature. Here we have identified 20 representative definitions that offer a more nuanced take on “fake news,” utilizing the various subdivisions in philosophy such as epistemology, philosophy of science, and ethics. In this subsection, much as in Section 2, we will go through the definitions offered and look for similarities between them. (At the same time, we will not seek to elaborate here on those categories that overlap sufficiently with those already encountered in Section 2.) We identify categories that the various definitions share in Table 2. These include the following: Mimicking, False, “Bullshit,” Lack of Concern for Truth, Intentional, and Social or Political Aspect. We will then turn to identifying the aspects of these categories related to rephrasing.

Table 2. Philosophy definitions of “fake news”: major tendencies

#	Author(s)	Mimicking	False	Bullshit	Lack of Concern for Truth	Intentional	Social or Political Aspect	Undefined
1.	Harris (2022)	X			X		X	
2.	Stewart (2021)							X
3.	Croce and Piazza (2021)	X			X	X	X	
4.	Dentith (2016)	X	X			X		
5.	Jaster and Lanius (2018)		X	X		X		
6.	Fallis and Mathiesen (2019)	X			X	X	X	
7.	Mukerji (2018)			X	X		X	
8.	Pepp et al. (2019)	X					X	
9.	Musi and Reed (2022)							X
10.	Anderau (2021)	X				X		
11.	Rini (2017)	X	X				X	
12.	Gelfert (2018)	X	X			X		
13.	Grundmann (2023)	X						
14.	Levy (2017)	X			X	X		
15.	McIntyre (2018)		X			X		
16.	Habgood-Coote (2019)							X
17.	Novaes and de Ridder (2021)							X
18.	Goldman and Baker (2019)		X			X		
19.	Galeotti (2019)	X	X	X			X	
20.	Ball (2021)	X			X		X	

The structure of the discussion of categories here proceeds as follows: first, we describe the data that we obtained by presenting the relevant statistical information, then we outline the elements of rephrasing discernible in the features of various definitions, and finally we look at some ethical aspects. The goal is to see how rephrasing manifests itself within these features. However, for the sake of space, we will not elaborate on those categories that overlap sufficiently with what was explored above in relation to computer science.

The Mimicking category is a significant element in the philosophical literature on “fake news.” In this category, we have included terms and phrases

such as “mimic,” “guise of news,” “news that represent itself as genuine,” “designed,” and “presented as genuine news.” We have discovered that the majority of definitions fall under this category: i.e. 12 of the 20 definitions (60%). The idea of mimicking is notable because it doesn’t necessarily rely on the truth value of its claims to be considered “fake news” (e.g. Harris 2022; Ball 2021; Galeotti 2019). The emphasis on mimicry underscores the deceptive strategy employed in “fake news”: by imitating the appearance and format of legitimate news outlets, “fake news” seeks to exploit the trust that audiences place in established journalistic practices. This mimicry makes it challenging for consumers to distinguish between authentic and false information.

In the context of “fake news,” rephrasing is instrumental in the process of mimicking genuine news sources (i.e. it is a subcategory of mimicking—mimicking can be an effect of rephrase, and rephrasing is one way to mimic). By carefully adjusting linguistic elements, creators of “fake news” replicate the style, tone and structural conventions of legitimate journalism to deceive readers. This mimicry is achieved through the strategic use of rephrasing, which allows for the alteration of authentic content or the fabrication of new content that closely resembles credible reporting. (There is an assumption involved here that mimicking rests upon, that will be elaborated on as potentially problematic in the next paragraph). Rephrasing enables the integration of falsehoods into a familiar journalistic framework, making the fabricated information appear trustworthy. This deliberate manipulation exploits the readers’ expectations and cognitive biases, particularly their trust in established news formats. As a result, the role of rephrasing in mimicking not only enhances the deceptive quality of “fake news,” but also poses significant challenges for detection and critical assessment. Recognizing the patterns of rephrased language that contribute to mimicry is therefore essential for developing analytical skills and critical thinking habits aimed at identifying and mitigating the impact of “fake news” in public discourse.

In order to understand the concept of mimicking, it is worth exploring the distinction between legitimate and illegitimate news sources, as the concept of mimicking in “fake news” has this distinction built in: after all, to mimic one must mimic something or someone. As was mentioned before, the category of mimicking rests on a potentially problematic assumption: it must assume that there is a difference between legitimate and illegitimate news sources. It is not always obvious when a news source is legitimate or not. This issue was touched upon by Habgood-Coote’s (2019) third argument about the propagandistic aspect of using the term “fake news.” The object

that the mimicker is emulating is the legitimate news source, and the mimicker is the illegitimate new source. This is not an insignificant assumption. The problem with the difference between legitimate and illegitimate new sources is captured in the following question: Who is the arbiter of what is legitimate or illegitimate? Let us suppose that we submit that the government is the arbiter of legitimacy. This might seem reasonable. However, intuition and historical experience tell us that we might well want to refer to some government-sourced news as “fake news.” For example, consider communist-run Poland (PRL) in the 1980s. Only government-sanctioned stories counted as legitimate news sources. The reporters would declare that the population was happy and that there was plenty of food in all of the shops. However, for many Polish people confronted by empty shelves this was evidently not the case. It was often reported that a common response from shopkeepers was just “*Nie ma*” [roughly: “it is not here”]. So do we believe the official news source, or our own eyes? This is a deeper problem, certainly worth exploring in terms of its potential implications; however, it lies outside of the scope of our paper.

Turning now to the category False, we find that this emerges as a less central one in philosophical definitions of “fake news,” appearing as it does in just 7 out of the 20 definitions analysed (35%). It encompasses terms such as “false,” “untrue,” “fabricated,” and “not supported by evidence,” highlighting the dissemination of information that deviates from factual accuracy (Dentith 2016; Gelfert 2018; McIntyre 2018; Galeotti 2019). The prevalence of falsehood in these definitions underscores the ethical and epistemological concerns associated with the spread of misinformation in public discourse. We take it that there is sufficient overlap with the category of “falsehood” in Section 2 to not warrant further elaboration.

The category of Bullshit, as articulated by Frankfurt (2005), appears in only 3 out of the 20 philosophical definitions analysed (15%). This category includes definitions where “fake news” is characterized by a lack of concern for the truth, with producers being indifferent to the veracity of the information they disseminate (see, e.g., Jaster and Lanius 2018; Muke-rji 2018). Unlike outright falsehoods or intentional lies, “bullshit” involves statements made without regard for their truthfulness, aiming instead to persuade or manipulate the audience for other purposes. (We ourselves only took into account papers that made explicit use of the Frankfurtian conception of “bullshit.”) The notion is significant because it shifts the focus from the content of the message to the attitude of the communicator towards truth. Producers of “fake news,” in this sense, are not necessarily committed to spreading falsehoods, but are indifferent to whether their

statements are true or false, as long as they achieve their desired effect. This indifference undermines the epistemic foundations of public discourse and poses challenges for identifying and addressing misinformation.

Rephrasing plays a significant role in the propagation of “bullshit” within “fake news.” The process of rephrasing allows communicators to construct messages that sound plausible and convincing while lacking substantive truth. In cases of “bullshit,” rephrasing is utilized to reshape information in a way that prioritizes rhetorical effectiveness over factual correctness. This involves the use of ambiguous language, rhetorical flourishes, or emotionally charged expressions that can mislead the audience by creating an impression of meaningful communication where there is none. For example, a statement such as “A significant number of experts agree that this policy will fail” can be rephrased as “A significant number of experts agree that this policy is doomed,” amplifying the persuasive force without providing concrete evidence or specifying sources.

The ethical implications of disseminating “bullshit” are also worth mentioning, as it contributes to a culture of indifference toward truth and undermines meaningful dialogue. When individuals or institutions prioritize rhetorical appeal over accuracy, they foster a communicative environment where truth is secondary to persuasiveness. This can lead to a normalization of deceptive communication practices, making it increasingly difficult for audiences to distinguish genuine information from manipulative rhetoric. Moreover, the spread of “bullshit” erodes public trust in information sources, which is especially damaging in contexts where informed decision-making is critical, such as politics or public health. The cumulative effect is a public discourse that is more susceptible to misinformation, where audiences may become cynical and disengaged, doubting all information regardless of its source.

The category Lack of Concern for Truth features prominently in philosophical definitions of “fake news,” appearing in 30% of the definitions analysed (6 out of 20). This category highlights an attitude where the producers of “fake news” exhibit indifference towards the veracity of the information they disseminate (e.g. Ball 2021; Levy 2017; Fallis and Mathiesen 2019). Unlike intentional deception, where falsehoods are propagated knowingly, a lack of concern for truth reflects a disregard for whether the information is true or false, so long as it serves the communicator’s purpose. This category is very similar to that of Bullshit. Here, we looked for terms other than “bullshit” that were used to express a lack of concern for truth. We take it there is sufficient overlap with the category of Bullshit (in Section 3) to not warrant further elaboration.

The Intentional category is featured prominently in philosophical definitions of “fake news,” appearing as it does in 9 out of the 20 definitions analysed (45%). This category emphasizes that the dissemination of “fake news” involves a deliberate intent to deceive or mislead the audience (see, e.g., Dentith 2016; Gelfert 2018; Rini 2017; Croce and Piazza 2021). The centrality of intentionality in these definitions underscores the ethical dimension of “fake news,” highlighting the purposeful actions of communicators who manipulate information to achieve specific objectives, such as influencing public opinion, advancing political agendas, or generating financial gain. We conclude that there is sufficient overlap here with the category of Intent and Purpose in Section 2 to not warrant further elaboration.

The category of Social or Political Aspect appears in 40% of the philosophical definitions analysed, featuring in 8 out of 20 cases. This one emphasizes that “fake news” often serves specific social or political purposes, such as manipulating public opinion, advancing political agendas, or undermining democratic processes (see, e.g., Harris 2022; Pepp et al. 2019; Rini 2017). The inclusion of social and political dimensions highlights the broader impact of “fake news” on society, and its potential to influence collective behaviours and attitudes. The social or political aspect underscores the fact that “fake news” is not merely an isolated communicative act, but is embedded within larger socio-political contexts. It often exploits existing social tensions, ideological divides, or political controversies to achieve its objectives. This dimension reflects the instrumental use of “fake news” as a tool for propaganda, disinformation campaigns, or social engineering. It would appear that there is sufficient overlap with the category of Harm in Section 2 to not warrant further elaboration.

The category of Undefined, meaning that no definition has been provided, can be observed in 20% of the philosophical literature analysed, with 4 out of 20 authors opting not to offer a formal definition of “fake news,” or arguing that seeking such a definition is futile. Notably, Habgood-Coote (2019) argues against having any such definition by contending that the term “fake news” is itself linguistically defective and propagandistic, since it lacks a stable and coherent meaning. Novaes and de Ridder (2021) follow Habgood-Coote’s (2019) lead and refrain from defining the term, arguing instead that the concept of “fake news” is either too ambiguous or too problematic to warrant a precise definition. This approach reflects a critical stance toward the term itself, suggesting that its usage may be more harmful than beneficial to public discourse. Musi and Reed (2022) sidestep the “fake news” debate by constructing a definition of “semi-fake news.” Stewart (2021) contends that “fake news” amounts to an umbrella

term that should be thought of as being composed of smaller parts: namely, “misinformation,” “disinformation,” and “misleading content.”

By examining the definitions of “fake news” in both computer science and philosophy, we have highlighted their key elements and their relevance to understanding rephrased misinformation. Computer science often emphasizes practical detection methods and the operational aspects of misinformation, focusing on falsehoods, intent, and their harmful impact. In contrast, philosophy provides a deeper exploration of intent, ethical considerations, and the broader social implications of “fake news.” Together, these perspectives allow us to better grasp the interplay between rephrased language and the construction of misleading narratives.

Rephrasing in “fake news” highlights deeper issues such as truth, ethics, and societal impact. A prominent feature is mimicry, where rephrased content imitates the style and format of legitimate journalism, exploiting public trust in traditional news sources. Philosophical definitions emphasize the role of intentionality, as rephrased “fake news” often reflects deliberate efforts to mislead, influence opinions, or serve specific political and social agendas. The concept of falsehood is central, as rephrasing introduces inaccuracies or distorts meaning while maintaining a facade of legitimacy. Philosophy also considers the phenomenon of indifference to truth, where creators prioritize rhetorical impact over factual accuracy, contributing to the erosion of public trust in media. Additionally, the social and political dimensions of rephrased fake news are critical, as such manipulations amplify ideological divisions and undermine constructive public discourse. These elements underscore the ethical and epistemological concerns associated with rephrased misinformation, offering valuable insights into its societal consequences and highlighting the need for critical engagement with deceptive communication practices.

With these theoretical frameworks in mind, we turn our attention to the linguistic study of rephrasing. By analysing how rephrasing operates as a powerful tool for misinformation, we aim to identify key cues and mechanisms that contribute to its influence. In the next section, we will delve into the linguistic research on rephrasing, exploring how it can be used both to deceive and to develop effective critical thinking strategies for identifying misinformation.

4. RESEARCH ON REPHRASE IN ARGUMENTATION THEORY AND CORPUS LINGUISTICS AS A FRAMEWORK FOR STUDYING “FAKE NEWS”

Taking into account the parallels identified in Sections 2 and 3 between the concept of “fake news” and that of rephrasing, we now turn to recent

research on rephrase, construed as furnishing a framework for exploring the potential of rephrase-analysis tools in the study of argumentative discourse. This exploration leads us to propose a research program centred on the deployment of models of rephrase when reformulating messages to help identify fake news.

It is worth pointing to some of the literature on rephrase. An account is given in Younis et al. (2023)—but in sum, one can say that the concept of “rephrase” is moving towards a definition. Rephrase is first found in Inference Anchoring Theory (IAT), as a propositional relation distinct from inference and conflict. A speaker’s intention is decisive: when two contributions are linked by rephrase, the second is meant as a restatement of the first that neither provides a reason for it nor opposes it; marking a segment as rephrase therefore rules out pro- or con-argument status. The authors then refine types, positing rephrase specification (the second contribution is narrower in meaning) and rephrase generalization (the second is broader), and noting that these operate along multiple semantic scales (quantitative, evaluative, part-whole) and appear to be among the most frequent subtypes in argumentative corpora. They contrast rephrase with neighbouring notions: reformulation (an “equivalence operation” at the semantic/pragmatic level) and paraphrase (approximate semantic equivalence). Because IAT defines rephrase by communicative function (non-inferential restatement) rather than semantic equivalence, there are reformulations that are not rephrases (e.g. “in other words” used to draw an inference), and rephrases that are not paraphrases (when meanings are not semantically close). The upshot is a functional, intention-sensitive account that distinguishes rephrase from both argumentative support/attack and from purely semantic sameness, while motivating a finer typology of subtypes.

One word about rephrase and intentionality. Rephrase may, but does not have to, be accompanied with a speaker’s intention to reformulate. Still, the intentional aspect of rephrasing may be the purposeful act of expressing a statement in different words, not merely to clarify, but also to achieve a strategic rhetorical effect, gain a persuasive advantage, gain deeper comprehension of an original message, emphasize a specific point, or manage the discourse flow. This intent may differentiate rhetorical uses of rephrase from mere paraphrasing or restatement, as it may carry a persuasive goal beyond pure reformulation. Thus, in recent rephrase studies (e.g. Konat et al, 2016; Younis et al, 2023) these intentions of rephrasing have been brought up for discussion.

Now with the background discussion on rephrase done, our exploration of the technique of rephrasing messages has emerged as a powerful yet

insidious tool in the dissemination of “fake news.” By altering language subtly, creators of misinformation can manipulate truth in ways that are both convincing and deceptive.

Given the tendencies towards defining “fake news” using certain key rephrase-related terms, such as “mimicking”⁴ and “fabricating,” as discussed in Sections 2 and 3, this section will explore key areas of study of rephrase with a view to developing future tools for identifying “fake news” in reformulated messages. To this end, we propose possible directions of future inquiry to supplement the existing critical thinking theories with a study of how rephrasing for the sake of generating and spreading “fake news” can mislead audiences, evade detection, and amplify the impact of such news. By synthesizing key features identified in computer science and philosophical discussions, we aim to illuminate the role of rephrasing in the construction and spread of misinformation. Furthermore, this section seeks to establish a foundation for using some critical thinking tools and analytical frameworks to recognize and counteract the deceptive nature of rephrased content. Through this exploration, we move closer to understanding how to mitigate the pervasive influence of fake news in contemporary discourse.

For the purpose of succinctly depicting the potential rephrase-related properties of communication typical of the dissemination of “fake news,” we list in Table 3 those properties that we have found to furnish a regularity in respect of definitions of “fake news” appearing in both computer science and philosophy.

As we observed in Section 2, the definitions in computer science include the following features of rephrasing in “fake news”:

- Using language to obscure or misrepresent the truth.
- Mimicking the original content to create deceptive yet credible outputs.

4. We are not trying to claim that mimicry is a subtype of rephrase, but rather that it is an effective linguistic tool for rephrase. Rephrase, in some use cases, such as misrepresenting an opponent’s statement in straw man fallacies, might instead be conceived as an instance of mimicking, in the sense of making a p' “pretend” to contain exactly the same content as an original statement p , whereas in fact a p' contains a modified content, which is accompanied by the intention of making the other party’s position easier to attack. In this respect, in some cases, mimicking can be an effect of rephrase. The goal of our paper, which is to explore the potential of rephrase studies in identifying some linguistic manifestations, among other key manifestations, was partly achieved through indicating and discussing mimicking as a discursive strategy that may (but does not have to) be achieved by means of rephrasing. As the exploration of this overlap has a potential for capturing those rephrase uses that may illicitly mimic original messages, we have given in our paper arguments in favour of the claim that, in this respect, the systematic study of rephrase may be incorporated into the broader toolset for identifying misinformation.

- Masking unverifiable claims by pretending to reference reliable sources.
- Exploiting linguistic and structural patterns to evade detection.

Table 3: Principal rephrase-related features of “fake news”: summary from Sections 2 and 3

Category	Feature
Mimicry	Imitation of the tone, style, and structure of legitimate sources to exploit trust
	Creating outputs similar to the original content but embedding distortions
Manipulation of Truth	Using language to subtly obscure or misrepresent the truth
	Introducing distortions while maintaining the appearance of legitimacy
Intentionality	Deliberate efforts to align rephrased content with specific agendas
	Often linked to operational motives (e.g. propaganda)
Verification and Credibility	Masking unverifiable claims through subtle linguistic adjustments
	Pretending to have credible sources while tweaking content
Emotional and Cognitive Impact	Provoking emotional responses such as fear or anger through rephrasing
Ethical and Epistemological Concerns	Erosion of public trust due to repeated exposure to rephrased misinformation
Social and Political Implications	Amplifying societal divisions through tailored rephrased narratives
Critical Thinking Relevance	Recognizing deceptive patterns through linguistic and pragmatic analysis

The definitions of “fake news” in philosophy, on the other hand, as discussed in Section 3, emphasize features relating to concepts such as—among others—trust, truth, factual accuracy, and knowledge:

- Mimicking legitimate formats to exploit trust in established institutions.
- Prioritizing rhetorical appeal over factual accuracy (manipulation of truth).
- Amplifying societal and political divisions through targeted rephrasing (social and political implications).
- Embedding nuanced distortions that subtly alter the original meaning.

Turning now to the main features of “fake news” as found in the definitions taken from the philosophical literature (see Section 3), let us point to the key features of fake news: (i) using language to *mislead or obscure the truth*; (ii) pretending to have credible sources of information; and (iii) ethical and social aspects.

When considering these features in terms of possible benefits of employing rephrase studies to identify these kinds of misinformation in discourse, we can observe that feature (i) of “fake news”—“obscuring the truth”—may have a great deal in common with misrepresenting the content of an original statement in cases of rephrase use. Applying this communication technique in discourse by means of the use of rephrase can be reinforced by using linguistic techniques of rephrasing. Once this is done, a rephrase output which looks very similar to an original input may in fact mimic the original content, and thus be employed as a subtle misinformation tool. Thus, the role of rephrase in generating and spreading misinformation may consist in modifying the linguistic surface of a message in such a way that the truth is made obscure.⁵ For example, consider the ever-updating news aggregate website the *Drudge Report*. This influential site, run by Matt Drudge, rephrases original news headlines in order to make them more attractive to his audience (Carr 2011).

Building on this idea, feature (ii)—“pretending to have credible sources of information”—further illustrates how rephrasing can be manipulated to mislead an audience. This tactic involves referring to objectively reliable sources while simultaneously tweaking the content or context of what those sources originally stated. This kind of rephrase use essentially consists of referring to a given source of information in an inadequate way by, e.g., not referring to the source directly, while misrepresenting what that source mentions. To a certain extent, this way of using rephrase is similar to performing those fallacious arguments from expert opinion that rely on manipulating the content of what a genuine authority has uttered. Likewise, in such cases, misuses of rephrasing can be intentionally employed to create misrepresented content which, if spread widely, can serve as efficient “fake news,” functioning effectively precisely because of the similarity of contents.

5. Here we have in mind such as issues as straw man fallacies (Visser et al. 2018): rephrase may, but does not have to, mislead audiences by making the news “fake”. We emphasize that rephrase, due to its linguistic and structural features, may be just one of the possible vehicles for misinforming people, and thus also for the spreading of “fake news.”

5. CONCLUSION AND FUTURE WORK

To conclude, we have conducted an investigation into how rephrasing contributes to the construction and persuasive power of “fake news” by analysing definitions from both computer science and philosophy. Our paper has:

- examined representative definitions to identify core rephrase-related features—such as mimicry, fabrication, intentionality and the strategic obscuring of truth—that underlie “fake news.”
- bridged a gap between the overlap between rephrase and “fake news” as exhibited in the literature in philosophy and in computer science, in order to propose specific areas of philosophical and linguistic research into rephrase, applied specifically to identify linguistic manifestations of certain instances of rephrase that may have the effect of “fake news.”

This approach, we think, advances the field by opening new research threads for scholars in discourse analysis, argumentation studies, and communication studies, and does so by examining the role played by rephrase in respect of identifying a new class of “fake news.” In contrast to some papers, such as that of Anderau (2021), whose work centres primarily on defining “fake news,” our study extends the inquiry by emphasizing the role of rephrasing as a dynamic communicative strategy. By doing so, we aim to provide insights that will hopefully prove useful for developing a more comprehensive framework linking theoretical discoveries with practical applications—thereby equipping scholars to explore how nuanced linguistic manipulations can be systematically detected and mitigated in public discourse.

The study of rephrase, far from being a marginal phenomenon in communication (Younis et al. 2023), offers valuable insights into a crucial mechanism for the spread of “fake news.” By examining how rephrased content subtly manipulates and reframes information, we can better understand its role in constructing persuasive yet misleading narratives. Rephrase serves as an effective model for studying and detecting a powerful class of “fake news” that thrives on the ambiguity created by linguistic shifts.

Both computer science and philosophy provide complementary approaches to understanding and combating rephrased “fake news.” In computer science, “fake news” is operationalized for detection, focusing on patterns, algorithms, and data-driven approaches to identifying misrepresentations. These methods aim to provide concrete, scalable tools for filtering and analysing information in digital environments. On the other hand, philosophy provides a deeper theoretical framework, focusing

on the ethical, epistemological, and communicative dimensions of “fake news.” Philosophical approaches emphasize the intent behind deception and the ways in which language, including rephrasing, can be weaponized to obscure truth and manipulate public perception.

These two fields together offer a robust framework for addressing the challenges posed by rephrase in “fake news.” While computer science provides the practical tools for detection, philosophy probes the deeper implications of how and why rephrasing manipulates understanding. Future interdisciplinary collaborations between these fields could enhance our ability to both detect and critically analyse “fake news,” fostering a more holistic approach to combating misinformation.

Furthermore, developing a comprehensive corpus of rephrase misuse and implementing rephrase-checking analytics could provide significant tools for combating the dissemination of manipulated content. These tools would not only enhance our capacity to detect misused rephrases but also highlight how rephrase impacts cognitive biases, potentially paving the way for broader applications in misinformation detection. Critical thinking plays a pivotal role in this context, offering a practical framework for individuals to build immunity against the subtle manipulations that rephrased “fake news” entails. By fostering analytical skills and a disposition for scepticism, we can empower individuals to better navigate the complex information ecosystems of today’s online platforms, where rephrased content often evades conventional detection systems.

Looking ahead, a comprehensive model aimed at capturing the nuances of rephrase within “fake news” strikes us as essential.⁶ Such a model would focus on identifying and classifying the specific linguistic strategies through which rephrasing subtly manipulates content, often making deceptive statements more persuasive and harder to detect. This would provide a theoretical framework that complements existing detection methods, particularly those grounded in computer science. The development of this model would not only enhance detection capabilities but also provide new ways to systematically categorize rephrased misinformation across different media platforms.

In addition, a large, annotated corpus dedicated to the study of rephrase misuse would certainly be beneficial in pursuit of this enterprise. This

6. We would like to note that the research is ongoing into the technical use of theoretical rephrase tools for designing multi-agent systems that employ rephrase and argumentation (see Uberta et al, forthcoming), where this is devoted to taking argumentation and rephrase studies as a theoretical model framework for AI tools to evaluate the effectiveness of those theories in argument and rephrase evaluation. Such work addresses the issue of the erroneous use of rephrase.

corpus would serve as a vital resource for future studies, enabling both qualitative and quantitative analysis of how rephrase is employed in misleading contexts. Such a resource would also support the development of rephrase-checking algorithms, which could be integrated into existing “fake news” detection systems, offering a more refined means of identifying subtle yet impactful manipulations. This corpus would also facilitate cross-disciplinary research, merging insights from linguistics, argumentation studies and computational analysis.

Another promising direction for future research lies in integrating insights from rephrase studies into practical tools for combating “fake news.” One such application could build upon existing systems like the Reason-Checking “fake news” app Evidence Toolkit (Visser et al. 2020), which focuses on identifying and countering misinformation. Expanding this toolkit to include rephrase analysis would allow for the detection of subtle linguistic manipulations that often evade traditional fact-checking methods. By systematically incorporating mechanisms to identify mimicked structures, adjusted tones, and rephrased content that obscures or distorts truth, such a tool could serve as a comprehensive resource for addressing the deceptive nature of rephrased “fake news.” This integration would bridge the gap between theoretical research on rephrasing and its practical implications, empowering users to identify and resist the influence of misinformation in digital communication.

Finally, future work should prioritise the integration of critical thinking instruction, specifically aimed at empowering individuals to recognise rephrasing strategies that contribute to the spread of “fake news.” Educational frameworks that incorporate critical thinking through practical exercises for detecting rephrased misinformation could offer a defence against “fake news” that evades traditional detection systems. This instruction should be tied directly to an analysis of how rephrasing strategies exacerbate societal polarization, particularly in digital communication environments. By addressing the intersection between rephrasing and “fake news,” these initiatives could ultimately contribute to reducing the divisive effects of misinformation in public discourse.

AUTHORS’ CONTRIBUTIONS

Mitchell Welle: Literature Review, Argument Development, Analysing the Definitions, Qualitative Analysis, Interpretation of Results, Formatting, Editing, Proofreading.

Marcin Koszowy: Framing the Problem, Conceptualization and Modelling, Argument Development, Qualitative Analysis, Editing.

BIBLIOGRAPHY

- Ajao, Oluwaseun, Deepayan Bhowmik, and Shahrzad Zargari. 2019. “Sentiment Aware Fake News Detection on Online Social Networks.” In *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8504–08. Piscataway, NJ: IEEE.
- Alam, Firoj, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimitar Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2022. “A Survey on Multimodal Disinformation Detection.” arXiv:2103.12541. <https://arxiv.org/abs/2103.12541>.
- Anderau, Glenn. 2021. “Defining Fake News.” *Kriterion—Journal of Philosophy* 35 (3): 197–215. <https://doi.org/10.1515/krt-2021-0019>.
- Ball, Brian. 2021. “Defeating Fake News: On Journalism, Knowledge, and Democracy.” *Moral Philosophy and Politics* 8 (1): 5–26. <https://doi.org/10.1515/mopp-2019-0033>.
- Berrondo-Otermin, Maialen, and Antonio Sarasa-Cabezuelo. 2023. “Application of Artificial Intelligence Techniques to Detect Fake News: A Review.” *Electronics* 12 (24): 5041. <https://doi.org/10.3390/electronics12245041>.
- Carr, David. 2011. “How Drudge Stayed on Top.” *New York Times*, May 15. <https://www.nytimes.com/2011/05/16/business/media/16carr.html>. Accessed November 5, 2025.
- Collins, Dave, and Pat Eaton-Robb. 2022. “Families Testify of Confrontations with Sandy Hook Deniers.” AP News, September 27. <https://apnews.com/article/shootings-texas-connecticut-alex-jones-waterbury-f3ef318375efc40dccc7067fb8f53e5e>. Accessed November 5, 2025.
- Conroy, Nadia K., Victoria L. Rubin, and Yimin Chen. 2015. “Automatic Deception Detection: Methods for Finding Fake News.” *Proceedings of the Association for Information Science and Technology* 52 (1): 1–4. <https://doi.org/10.1002/pra2.2015.145052010082>.
- Croce, Michel, and Tommaso Piazza. 2021. “Misinformation and Intentional Deception: A Novel Account of Fake News.” In *Virtues, Democracy, and Online Media: Ethical and Epistemic Issues*, edited by Maria Silvia Vaccarezza and Nancy E. Snow, 52–67. New York: Routledge.
- Dentith, M.R.X. 2016. “The Problem of Fake News.” *Public Reason* 8 (1–2): 65–79.
- Desamsetti, Sankar, Satya Hemalatha Juttuka, Yamini Mahitha Posina, S. Rama Sree, and B.S. Kiruthika Devi. 2023. “Artificial Intelligence Based Fake News Detection Techniques.” In *Recent Developments in Electronics and Communication Systems*, edited by K.V.S. Ram Ramachandra Murthy, Sanjeev Kumar, and Mahesh Kumar Singh, 374–80. *Advances in Transdisciplinary Engineering* 32. Amsterdam: IOS Press. <https://doi.org/10.3233/ATDE221284>.
- Diaz-Garcia, Jose A., Dolores Ruiz M., and Martin-Bautista Maria J. 2025. “Liars Know How to Argue: An Approach to Disinformation Analysis Based on Argument Mining.” In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, vol. 3, edited by Marie-Jeanne Lesot, Susana Vieira, Marek Z. Reformat, João Paulo Carvalho, Fernando Batista, Bernadette Bouchon-Meunier, and Ronald R. Yager, 60–69. Cham: Springer.
- Dong, Yiqi, Dongxiao He, Xiaobao Wang, Yawen Li, Xiaowen Su, Di Jin. 2023. “A Generalized Deep Markov Random Fields Framework for Fake News Detection.” In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, edited by Edith Elkind, 4758–65. International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2023/529>.
- Durden, Tyler. 2024. “US Women’s Life Expectancy Is the Lowest among Developed Nations.” ZeroHedge, September 14. <https://www.zerohedge.com/medical/us-womens-life-expectancy-lowest-among-developed-nations>. Accessed November 5, 2025.

- Fallis, Don, and Kay Mathiesen. 2019. "Fake News Is Counterfeit News." *Inquiry* 1–20. <https://doi.org/10.1080/0020174X.2019.1688179>.
- Farajtabar, Mehrdad, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, and Hongyuan Zha. 2017. "Fake News Mitigation via Point Process Based Intervention." arXiv:1703.07823. <http://arxiv.org/abs/1703.07823>.
- Fifita, Faizi, Jordan Smith, Melissa B. Hanzsek-Brill, Xiaoyin Li, and Mengshi Zhou. 2023. "Machine Learning-Based Identifications of COVID-19 Fake News Using Biomedical Information Extraction." *Big Data and Cognitive Computing* 7 (1): 46. <https://doi.org/10.3390/bdcc7010046>.
- Frankfurt, Harry G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- Freiberg, Nava. 2025. "Netanyahu acknowledges Israel losing online 'propaganda war,' should be doing more." *Times of Israel*, August 10. https://www.timesofisrael.com/liveblog_entry/netanyahu-acknowledges-israel-losing-online-propaganda-war-should-be-doing-more/. Accessed November 5, 2025.
- Galeotti, Elisabetta. 2019. "Believing Fake News." In *Post-Truth, Philosophy and Law*, edited by Angela Condello and Tiziano Toracca, 37–52. London: Routledge.
- Gelfert, Axel. 2018. "Fake News: A Definition." *Informal Logic* 38 (1): 84–117. <https://doi.org/10.22329/il.v38i1.5068>.
- Goldman I, Baker D (2019) Free speech, fake news, and democracy. *First Amendment Review* 18(1): 66–145
- Grundmann T (2023) Fake news: the case for a purely consumer-oriented explication. *Inquiry* 66(10):1758–1772, DOI 10.1080/0020174X.2020.1813195, URL <https://doi.org/10.1080/0020174X.2020.1813195>, <https://doi.org/10.1080/0020174X.2020.1813195>
- Habgood-Coote, Joshua. 2019. "Stop Talking about Fake News!" *Inquiry* 62 (9–10): 1033–65. <https://doi.org/10.1080/0020174X.2018.1508363>.
- Harris, Keith Raymond. 2022. "Real Fakes: The Epistemology of Online Misinformation." *Philosophy & Technology* 35 (3): 1–24. <https://doi.org/10.1007/s13347-022-00581-9>.
- Hassan, Aumyo, and Sarah J. Barber. 2021. "The Effects of Repetition Frequency on the Illusory Truth Effect." *Cognitive Research: Principles and Implications* 6 (38): 1–12. <https://doi.org/10.1186/s41235-021-00301-5>.
- Jain A, Kasbe A (2018) Fake news detection. In: 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp 1–5, DOI10.1109/SCEECS.2018.8546944
- Jaster, Romy, and David Lanius. 2018. "What Is Fake News?" *Versus* 2 (127): 207–27.
- Kandell J (2024) Daniel c. dennett, widely read and fiercely debated philosopher, 82, dies. URL <https://www.nytimes.com/2024/04/19/books/daniel-dennett-dead.html>
- Khattar, Dhruv, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. "MVAE: Multimodal Variational Autoencoder for Fake News Detection." In *Proceedings of the 2019 World Wide Web Conference*, 2915–21. New York: ACM.
- Konat, Barbara, Katarzyna Budzynska, and Patrick Saint-Dizier. 2016. "Rephrase in Argument Structure." In *Foundations of the Language of Argumentation: COMMA 2016 Workshop*, 32–39. Potsdam: University of Potsdam.
- Koszowy, Marcin, Steve Oswald, Katarzyna Budzynska, Barbara Konat, and Pascal Gygas. 2022. "A Pragmatic Account of Rephrase in Argumentation: Linguistic and Cognitive Evidence." *Informal Logic* 42 (1): 49–82. <https://doi.org/10.22329/il.v42i1.7212>.
- Kou Z, Shang L, Zhang Y, Yue Z, Zeng H, Wang D (2022) Crowd, expert ai: A human-ai interactive approach towards natural language explanation-based covid-19 misinformation detection. International Joint Conferences on Artificial Intelligence Organization

- Levy, Neil. 2017. “The Bad News about Fake News.” *Social Epistemology Review and Reply Collective* 6 (8): 20–36.
- Major, Mark, and David J. Andersen. 2016. “Polls and Elections: Swift Boating Reconsidered: News Coverage of Negative Presidential Ads.” *Presidential Studies Quarterly* 46 (4): 891–910. <https://doi.org/10.1111/psq.12324>.
- McIntyre, Lee C. 2018. *Post-Truth*. Cambridge, MA: MIT Press.
- Mukerji, Nikil. 2018. “What Is Fake News?” *Ergo: An Open Access Journal of Philosophy* 5: 923–946. <https://doi.org/10.3998/ergo.12405314.0005.035>.
- Musi, Elena, and Chris Reed. 2022. “From Fallacies to Semi-Fake News: Improving the Identification of Misinformation Triggers across Digital Media.” *Discourse & Society* 33 (3): 349–70. <https://doi.org/10.1177/09579265221076609>.
- Nakov P, Corney D, Hasanain M, Alam F, Elsayed T, Barr’ on-Cede~no A, Papotti P, Shaar S, Martino GDS (2021) Automated fact-checking for assisting human fact-checkers. URL <https://arxiv.org/abs/2103.07769>, 2103.07769
- “Netanyahu Declares War on Free Speech as Israel’s Propaganda Efforts Falter.” 2025. *InfoWars*, September 4. <https://www.infowars.com/posts/netanyahu-declares-war-on-free-speech-as-israels-propaganda-efforts-falter>. Accessed November 5, 2025.
- Novaes, Catarina Dutilh, and Jeroen de Ridder. 2021. “Is Fake News Old News?” In *The Epistemology of Fake News*, edited by Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann, 155–78. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198863977.003.0008>.
- Oshikawa R, Qian J, Wang WY (2018) A survey on natural language processing for fake news detection. CoRR abs/1811.00770, URL <http://arxiv.org/abs/1811.00770>, 1811.00770
- Parikh, Shivam B., and Pradeep K. Atrey. 2018. “Media-Rich Fake News Detection: A Survey.” In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 436–441. <https://doi.org/10.1109/MIPR.2018.00093>.
- Pepp, Jessica, Eliot Michaelson, and Rachel K. Sterken. 2019. “What’s New about Fake News?” *Journal of Ethics and Social Philosophy* 16 (2): 67–94. <https://doi.org/10.26556/jesp.v16i2.629>.
- Perez-Rosas, Verónica, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. “Automatic Detection of Fake News.” arXiv:1708.07104. <http://arxiv.org/abs/1708.07104>.
- Qian, Feng, Chengyue Gong, Karishma Sharma, and Yan Liu. 2018. “Neural User Response Generator: Fake News Detection with Collective User Intelligence.” In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 3834–40. International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2018/533>.
- Rini, Regina. 2017. “Fake News and Partisan Epistemology.” *Kennedy Institute of Ethics Journal* 27 (S2): 43–64. <https://doi.org/10.1353/ken.2017.0025>.
- Ruchansky, Natali, Sungyong Seo, and Yan Liu. 2017. “CSI: A Hybrid Deep Model for Fake News Detection.” arXiv:1703.06959. <https://arxiv.org/abs/1703.06959>.
- Sharma, Upasna, and Jaswinder Singh. 2024. “A Comprehensive Overview of Fake News Detection on Social Networks.” *Social Network Analysis and Mining* 14 (1): 1–20. <https://doi.org/10.1007/s13278-024-01280-3>.
- Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. URL <https://arxiv.org/abs/1708.01967>, 1708.01967
- Singhal, Shivangi, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnuram Kumaraguru, and Shin’ichi Satoh. 2019. “SpotFake: A Multimodal Framework for Fake News Detection.” In

- 2019 *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 39–47. <https://doi.org/10.1109/BigMM.2019.00-44>.
- Stewart, Elizabeth. 2021. “Detecting Fake News: Two Problems for Content Moderation.” *Philosophy & Technology* 34 (4): 923–940. <https://doi.org/10.1007/s13347-021-00442-x>.
- Tschiatschek, Sebastian, Adish Singla, Manuel Gomez-Rodriguez, Arpit Merchant, and Andreas Krause. 2017. “Fake News Detection in Social Networks via Crowd Signals.” arXiv:1711.09025. <http://arxiv.org/abs/1711.09025>.
- Uberna, Maciej, Wawer, Michał, Chudziak, Jarosław and Koszowy, Marcin. Forthcoming. “Rephrasing Agents: Theoretically Driven Multi-Agent System for Identifying Misuses of Reformulation.” Under review.
- UN. 2014. “Country Classifications.” https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf. Accessed November 5, 2025.
- Visser, Jacky, John Lawrence, and Chris Reed. 2020. “Reason-Checking Fake News.” *Communications of the ACM* 63 (11): 38–40. <https://doi.org/10.1145/3397189>.
- Visser, Jacky, Marcin Koszowy, Barbara Konat, Kasia Budzynska, and Chris Reed. 2018. “Straw Man as Misuse of Rephrase.” In *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation*, vol. 2, edited by Steve Oswald and Didier Maillat, 941–62. London: College Publications.
- Von der Leyen, Ursula. 2024. “Special Address by Ursula von der Leyen, President of the European Commission.” Speech presented at the World Economic Forum Annual Meeting, Davos, January 16. <https://www.weforum.org/meetings/world-economic-forum-annual-meeting-2024/sessions/special-address-by-ursula-von-der-leyen-president-of-the-european-commission-96293a5a9d/>. Accessed November 5, 2025.
- Wang, Yaqing, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. “EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection.” In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 849–57. New York: ACM. <https://doi.org/10.1145/3219819.3219903>.
- WHO. 2024. “Slovakia Health Data Overview for the Slovak Republic.” World Health Organization. <https://data.who.int/countries/703>. Accessed November 5, 2025.
- Wu L, Rao Y (2020) Adaptive interaction fusion networks for fake news detection. CoRR abs/2004.10009, URL <https://arxiv.org/abs/2004.10009>, 2004.10009.
- Younis, Ramy, Daniel de Oliveira Fernandes, Pascal Gyax, Marcin Koszowy, and Steve Oswald. 2023. “Rephrasing Is Not Arguing, but It Is Still Persuasive: An Experimental Approach to Perlocutionary Effects of Rephrase.” *Journal of Pragmatics* 210: 12–23. <https://doi.org/10.1016/j.pragma.2023.03.010>.
- Zhang X, Ghorbani AA (2020) An overview of online fake news: Characterization, detection, and discussion. *Information Processing Management* 57(2):102025, DOI <https://doi.org/10.1016/j.ipm.2019.03.004>, URL <https://www.sciencedirect.com/science/article/pii/S0306457318306794>
- Zhou, Xinyi, and Reza Zafarani. 2018. “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities.” arXiv:1812.00315. <http://arxiv.org/abs/1812.00315>.
- Zhou, Xinyi, Jindi Wu, and Reza Zafarani. 2020. “SAFE: Similarity-Aware Multi-Modal Fake News Detection.” arXiv:2003.04981. <https://arxiv.org/abs/2003.04981>.