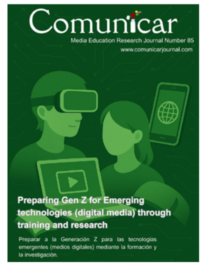
Comunicar, n. 85, v. XXXIV, 2026 | Media Education Research Journal | ISSN: 1134-3478; e-ISSN: 1988-3478
www.comunicarjournal.com



Mind the Gap: "Traditional" vs. Computational Research Logics in Fake News Detection

La brecha: Lógicas «tradicionales» y computacionales en la detección de noticias falsas

Svetlana Bodrunova, St. Petersburg State University, St. Petersburg (Russia) (s.bodrunova@spbu.ru) (https://orcid.org/0000-0003-0740-561X)

ABSTRACT

Fake news detection has become an acutely important goal in both academic studies and editorial practice, creating a research area that comprises journalistic debunking of fakes, cross-disciplinary fact-checking projects, and automated efforts of fake news detection. However, with the growth of these industries, an epistemological gap between "traditional" (conceptual, qualitative, quantitative, or mixed-methods) and computational studies of detecting fakes has been deepening. We describe the two divergent logics of fake definition and detection. In particular, international regulation, industrial fake detection, and most media studies legitimize the "blurred border" between fact and interpretation, warning against too strict elimination of fakes and preserving freedom of expression. Computational methods, in their turn, are better in automated fake detection but are "yes/no"-oriented, often ignoring the variety of interpretive forms in public communication. Rooted deeper than just in individual research designs, the divergence of logics sharpens when the public need of clear-cut fake detection runs into freedom of interpretation that results from centuries of struggle for standards in public speech and journalism. Employing critical reviewing of 45 conceptual academic and industrial writings, we outline the major shortcomings of the lack of clear textual markers for fake news in "traditional" media studies, on one hand, and of the "yes/no" logic in computational fake detection, on the other. We propose a five-pillar epistemological framework for fake detection, including true/false, fact/interpretation, discrepancy/solidity, media/user-generated evidence, and human/AI authorship dimensions.

RESUMEN

La detección de noticias falsas se ha convertido en un objetivo de suma importancia tanto en los estudios académicos como en la práctica editorial, creando un área de investigación que comprende la desacreditación periodística de falsificaciones, proyectos interdisciplinarios de verificación de hechos y esfuerzos automatizados de detección de noticias falsas. Sin embargo, con el crecimiento de estas industrias, se ha profundizado una brecha epistemológica entre los estudios «tradicionales» (conceptuales, cualitativos, cuantitativos o mixtos) y computacionales para detectar falsificaciones. Describimos las dos lógicas divergentes de definición y detección falsas. En particular, la regulación internacional, la detección de falsificaciones industriales y la mayoría de los estudios de medios legitiman la «frontera borrosa» entre los hechos y la interpretación, advirtiendo contra la eliminación demasiado estricta de las falsificaciones y preservando la libertad de expresión. Los métodos computacionales, a su vez, son mejores en la detección automática de falsificaciones, pero están orientados al «sí/no», ignorando a menudo la variedad de formas interpretativas en la comunicación pública. Arraigada más profundamente que solo en diseños de investigación individuales, la divergencia de lógicas se agudiza cuando la necesidad pública de una detección falsa clara se encuentra con la libertad de interpretación que resulta de siglos de lucha por los estándares en el discurso público y el periodismo. Empleando una revisión crítica de 45 escritos conceptuales académicos e industriales, describimos las principales deficiencias de la falta de marcadores textuales claros para las noticias falsas en los estudios de medios «tradicionales», por un lado, y de la lógica «sí/no» en la detección computacional de noticias falsas, por el otro. Proponemos un marco epistemológico de cinco pilares para la detección falsa, que incluye verdadero/falso, hecho/interpretación, discrepancia/solidez, evidencia generada por los medios/usuarios y dimensiones de autoría humana/IA.

KEYWORDS | PALABRAS CLAVE

Fake News, Detection of Fakes, Computational Communication Studies, Conceptual Model, Conceptual Gap, Epistemology.
Noticias falsas, detección de falsificaciones, estudios de comunicación computacional, modelo conceptual, brecha conceptual, epistemología.

Received: 2025-08-18 | Reviewed: 2025-10-31 | Accepted: 2025-11-02 | Online First: 2026-04-11 | Published: 2026-04-15
DOI: https://doi.org/10.5281/zenodo.19690365 | Pages: 101-113

1. Introduction

Authenticity has for decades been perceived as one of the foundational trust-building mechanisms in various communication fields and professions, from interpersonal dialogue to journalism, PR, and advertising, even if its functions differed depending of the sub-field. However, in the recent decades, information that distorts or denies the genuine construction of truth as a ground for mutual understanding has moved from individual-case to mass-scale level, which has received a popular name of "the post-truth epoch." The reaction of the academe to the rise of non-authenticity in public communication has been equally massive; a multitude of approaches to understanding, detecting, and battling what is known as fake news have appeared.

One of today's distinct gaps in communication science is low connection between "traditional" (theories-driven and non-computational in methods) and data-driven/computational approaches to many communication phenomena. Fake/false news is no exception, as "traditional" and computational approaches use differing grounds for definition and detection of fake news. The discursive field that discusses fakes is already highly diverse and both benefits and suffers from this diversity, which negatively affects academic efficiency in fakes detection and industrial fact-checking strategies. However, the conceptual grounds that make traditional and computational logics of fake detection diverge more and more have not yet been substantially addressed; thus, the research goals of our paper are (1) to define the essence of this divergence and (2) to propose the remedies for narrowing the logical gap between the two branches of communication science.

In this paper, we focus on a major conceptual gap in fake news studies, which is the one between "traditional" media and communication studies of falsehood in news, on one hand, and computational communication studies of fake news, on the other. The former and the latter both have their advantages and disadvantages in how they conceptualize fakes; but the biggest issue, to our viewpoint, is their low interconnectedness and divergence in logics.

Therefore, our review is conceptual in its nature, and thus its methodology is informed by both the strict PRISMA rules (Tricco et al., 2018; updated in 2025) and the approaches to writing conceptual papers (Jaakkola, 2020; Reese, 2023). We are aiming at reviewing influential and highly cited works that in certain ways discuss the (dis)advantages of the existing approaches and do not wish to review all the works that propose particular methods of fakes detection.

The paper selection process was based on search within three major databases that, taken together, cover English-language works in social science, engineering, and computer science (EBSCO, IEEE, and Academic Search Ultimate). The additional sources used for checking the indexation and popularity of the papers were Google Scholar, Scopus, and Web of Science Core Collection databases. Additionally, the sampling included the academic papers dedicated to conceptual description of the major industrial fake detection projects. The initial sample included 746 papers. However, most of them described individual attempts of fakes detection. We eliminated the papers that did not contain attempts of conceptualization on the nature of fake information, as well as systemic reviews of the empirical studies. This has left us with 137 papers/chapters of more conceptual stance which dealt with conceptual classifications of fakes, their historic and policy analysis, and theorizing falsehood. Of them, 39 papers were selected based on the source of publication, scholarly attention, and consistency of argument; six more papers were added to this list, since, as stated above, they were conceptually describing industrial attempts to create multi-class fake classifiers.

After preliminary reading of the 137 papers of the reduced sample, we have hypothesized on the nature of the gap between traditional and computational communication studies with the regard to fakes detection. In particular, we argue that non-computational media and communication studies have in focus the hazardously (but necessarily, as we show below) blurred border between clearly (and unacceptably) wrong information and various forms of fluctuation of meaning between fact and interpretation, from bias to satire. They mostly accept that information authenticity distortions may be widely ranged and have no clear-cut yes/no belonging to the realms of either true or false information. Simultaneously, they offer a multitude of approaches to what fake news is, how it is used, and why it is used (Waisbord, 2018) in the absence of "commonly agreed typology framework, specific categorization criteria, and explicit definitions" (Kapantai et al., 2021: 1). Such multitude of approaches, definitions and understandings of misinformation, disinformation, false information and fake news (which are sometimes used interchangeably while they should not) has created conflicting conceptualizations, followed by calls for conceptual clarity (Khan et al., 2022; Simons & Manoilo, 2021) and rare enough interdisciplinary efforts (Lazer et al., 2018). This conceptual abundance prevents, at least partly, provision of widely shared methodologies of fake detection that would bear in mind the thin division lines between fake, information distortion, and legitimate interpretation. Computational communication studies (Hilbert et al., 2019; van Atteveldt & Peng, 2018), in their turn, have been better at providing means of fake detection, but this, in most cases, has demanded the conceptual reduction of what fakes are to the "yes/no" binary logic. The latter regrettably excludes the "blurred border" idea of gradations in interpretation of truth, upon the importance of which international regulators and the EU authorities clearly insist. This difference between the two areas of fakes studies arises partly due to the fact that the "traditional" zone is more theory-driven in its nature, while the computational zone is more data-driven and overall relies more on methods as its foundation, rather than on theories and previous knowledge (Strasser, 2012).

Below, we convey our argument developed during and after closer manual analysis of the 45 papers of the final sample. The remainder of the paper is organized as follows. We first describe the two divergent logics of traditional and computational communication studies of fakes, highlighting the professional, conceptual, and research-related premises for formation of such logics, as well as their advantages and shortcomings. Then we propose a five-pillar epistemological framework that aims at bridging the gap between traditional and computational approaches, also suggesting that the framework may serve as a pipeline for developing a closer-to-human-logic multi-class fake detector. In the conclusion, we discuss implications of cross-pollination between the two branches of fake detection research, as well as some recommendations for future detection of false information.

2. The "Blurred Border": Definitions and Typologies of Fakes in Media Studies

Among others, one key issue in academic discussion on fake news seems to be of concern for the media and communication studies community. This issue is that of gradations of authenticity as closeness to objective truth. It arises due to a paradox of well-felt but still blurred boundaries between intentionally false informing and distorted information in a wider sense. The current UN and EU legal approaches to information disorders (Baade, 2018; Tumber & Waisbord, 2021; Wardle & Derakhshan, 2017) are based on the normative call for distinguishing between strictly and intentionally false informing ("disinformation"), on one hand, and distorted information, on the other. False news needs to be countered, while distorted news needs to be legally tolerated in order to avoid an even worse abuse of free speech and freedom of information in attempts to counter interpretations. Thus, the border between true and false information is seen as fundamentally (and acceptably) blurred, but a search of ways to distinguish disinformation from other types of fictitious speech is a goal.

This distinction is partly mirrored in "traditional" media studies of fakes that search for epistemic, historic, and genre-related grounds of fakes, as well as link fakes to actors of communication, like editorial offices, audiences, or news sources. In contrast to this, computational communication studies of fakes, as well as efforts by the industry giants, successfully enough seek to find textual grounds for automated fake news detection, but often see the fake/non-fake classification in a rather binary way. The two research areas rarely use each other's achievements, which calls for their intersection and collaboration. Their difference is, of course, speculative, fuzzy, and non-absolute; but it serves well for discovering the conceptual conflicting gaps they both face.

A major gap in "traditional" media and communication studies of fakes is that they lack clear and agreed-upon textual markers of fakes that would correspond to various types of information disorders. Below, we reconstruct some approaches that point out to such necessary but still non-well-defined textual markers.

If one starts from definitions, we need to note that not all the media/social-science papers on fake news problematize the "blurred border" issue when defining fake/false news. For example, a group of academic leaders has provided a most general definition for fake news as "fabricated information that mimics news media content in form but not in organizational process or intent" (Lazer et al., 2018: 1094). Despite the elegance of this definition published in the *Science* policy-oriented sub-journal *Insights*, it allows for neither clearly detecting fakes (as we can rarely provenly discover the organizational process or intent from a piece of news, and this is exactly where the problem with fakes lies!) nor delineating dubious interpretations from false facts. However, this definition may point out to what lacks in the fake-news texts: The metatextual signs of editorial judgment, news production process, and journalistic intention.

On all of these, perhaps, the media community needs to become more explicit, to make the news it produces positively more distinguishable from fakes. These elements may, in future, allow for better detecting the metatextual lacunas in fake news, both by the readership and automated filters. If fakes "claim to be actual" (Simons & Manoilo, 2021: 36) by mocking genre elements of news, professional metatext needs to start playing a fake/non-fake gatekeeper role, additionally linking the content features to the verifiable and reputable news producers, as well as distinguishing between false informing and (arguably) legitimate structural biases in news.

If not in definitions, in media studies of fakes, the "blurred border" idea shimmers via ubiquitous mentions of "mis- and disinformation." Such mentions indicate that the regulators' framework has been generally accepted. And yet, surprisingly rare academic works, and even more rare fact-checking initiatives, operationalize this particular EU distinction in research designs to detect various classes of fakes. This may be because, despite the seeming clarity, the "misinformation vs. disinformation" distinction does not yet provide for clear textual markers when it comes to finding fakes, while it could. The discursive difference between misinformation and disinformation is not well addressed in today's scholarly conceptualization of fake news.

In addition, the mental associations that shape the "fake news" thinking tend to be gradually shifting from satire-like representations of reality to the focus upon deceptive information (Tandoc Jr et al., 2018a), with attempts to abuse the term by also calling politically oppositional information fake news (Simons & Manoilo, 2021; Waisbord, 2018). This makes conceptualizations of fakes and "constructs related to fake news" (Khan et al., 2022: 3) vary also in time, but does not help in coming closer to conceptualizing how exactly intentional disinformation differs from other types of information disorders. Even an evident distinction between the satire coming from *The Onion* or *The Daily Currant* and fake news noted, e.g., by Berghel (2017) is not yet well-described in textual terms.

In search for responses to these challenges, in several research and review papers, the "blurred border" idea has transformed into typologies of fakes, based on assessment of normative text features traditional for journalism. Of such normative features, facticity/deceptiveness, author intention, and (potential) harm are most commonly used by both scholars and industrial players (Kapantai et al., 2021; Tandoc Jr et al., 2018a), even if the (potential) harm relates more to effects of a fake text than to its nature.

Of the three normative text features, the most problematic in epistemic terms is the author intention, as judgment upon it is, at least partly, imposed by the reader/viewer. Despite that, taking the author intention into account when both theoretically conceptualizing and practically detecting fakes may help cut news satire, parodies, or advertising from deliberation-oriented falsehoods like fabrication of texts and videos or systemic propaganda. However, strongly proving malicious intent (and which one in particular, and whose one exactly), not nature or effect, behind a single pseudo-journalistic text or a social media post may usually go beyond the scope of academic inquiry, while being a major focus of attention of social media giants. When given to the mercy of social media platforms only, detecting intent may also quickly breach the "blurred border" stance of reasonable tolerance as understood by regulators.

At the same time, using the term "online falsehoods" that would comprise fakes, dis-, and misinformation, Lien et al. (2022) suggest that tech platforms are not unanimous in their reactions to such falsehoods and use a range of collaborative, monetary, and restrictive counter-actions, from supporting third-party fact-checkers to content moderation (p. 8) – where, we would add, both practices are not particularly transparent in terms of markers of fakes. Most interventions of tech companies targeting online falsehoods fluctuate between the declared corporate social responsibility / serving the community and their brand damage repair / response to the blame discourse. It forms the platforms' understanding of fake news as a clearly applied one, stemming from community conventions, on one hand, but dependent on arbitrary decisions by one actor (a platform), on the other.

Here, again, a need for conceptual and operational (that is, textual) gradation of information disorders is evident, just as it is necessary to de-politicize, substantiate, and diversify the term in the vernacular use, both by elites and citizens (Khan et al., 2022), as well as by platforms. Such depoliticization is especially needed for less democratic contexts. Hopefully, clearer understanding of what cannot and should not be recognized as a fake news may deprive autocratic leaders of the excuses to oppress opposition by calling oppositional bias fakes, extremism, or sensitive information. And, wider, we need to constantly question from whom, even in democracies, the "information order" and "information disorder" understandings come (say, political parties, dominant social groups, populist leaders, media corporations, or IT giants), as defining fakes may be seen as a potential mechanism for exclusion and discrimination. And it is of particular importance whether and how exactly the primary definers of the discourse on fake news see (if ever) the textual shape of fakes – and what lacks in their descriptions of falsehoods.

And, no doubt, a major question is whether professional journalists and audiences can clearly distinguish the intolerable and undoubtedly harmful fake news from other questionable forms of public dialogue that need to be fought by public consensus, not by law or algorithmic means. Thus, Nielsen and Graves (2017) demonstrated that other types of news distortions like extreme bias, propaganda, or advertorials were more efficiently detected by audiences than fakes masquerading as news. Evidently, the spread of fake news has also brought new challenges for journalists themselves, questioning their degree of responsibility and accountability for fighting misinformation and fakes, as well as their own ability to distinguish authentic news from falsified ones and the professional journalistic tools for such disambiguation, including textual ones.

Saldaña and Vu (2022) showed that journalists cared about fake news and about the effects false information might have on journalism, news organizations, democracy, and audiences. Interestingly, for journalists more concerned about fakes news, the detection mechanisms work better: Regardless the revealed low level of confronting and reporting fake news, Saldaña and Vu's study has identified that the "sense of reciprocity toward social media audiences <and> opinion of the role social media companies play in the news ecosystem" significantly increase the likelihood of journalists' debunking misinformation (p. 14). Schapals (2018) has demonstrated that journalists are interested in understanding better how misinformation spreads online and in stronger collaboration with independent initiatives that fight fakes; however, little evidence had been there on how exactly journalists range fakes and work with the "blurred border" in fake detection. In this respect, an important area is contextualization of fake news recognition – that is, the study of differences in journalistic practices of fake debunking in countries with different degrees of journalistic professionalism, as well as those of trust in the news media.

Humprecht (2020) in this vein looked into how these two factors – professionalism and trust – affect the level of source transparency in different national and cultural contexts, showing that "the use of source transparency is linked to characteristics of the information environment and to the type of fact-checking organization" (p. 13), thus bridging the textual, normative, and organizational aspects of fake recognition. Furthermore, scholars showed that while, in the American context, fact-checking is mainly a journalistic practice, globally there is more diversity (Graves, 2018), so more diversity may lie at the heart of marking news as fake.

The same goes for the audience, as fake news "is in the eye of the beholder" (Simons & Manoilo, 2021: 49). While news is constructed by journalists, fake news is co-constructed by the audience. As Nielsen and Graves (2017: 1) showed, audiences tend to use exactly the "blurred border" logic in how they conceptualize fakes: "People see the difference between fake news and news as one of degree rather than a clear distinction."

In the constructivist manner, fakeness of consumed information depends highly on whether the audiences perceive the fake as real (Tandoc Jr et al., 2018b), which is one of the major foci in media studies of fake news. Among other works, Wagner and Boczkowski (2019) looked at the practices and interpretations around the perception of misinformation consumption in today's USA, showing that people were concerned about issues of trustworthiness and quality on social media platforms, dissatisfied with the quality of news reporting, but perceived fake news more as a non-faceted and self-explanatory concept. Thus, by contributing to spreading and interpreting fakes, the audiences curate fake news streams (Waisbord, 2018) – often without proper understanding of what fake news is and most surely without being able to recognize it via markers.

The scholars understandably try to clarify what lies behind users' ability to recognize fakes, and cognitive studies on fake recognition seem to have come the closest to assessing the combinations of textual features, background knowledge, and cognitive/psychological personality traits that shape ability of individuals or groups to detect fakes – as Zhang and Ghorbani (2020: 1) have put it, "the user, content, and context." Here, two strategies exist that, eventually, tackle similar constellations of cognitive, textual, and background factors of fake detection, but the first approach induces and tests them experimentally while the second one deduces the factors from surveys. Within the former approach, Porshnev and colleagues (2021) have shown that conspiracy-theory-oriented styles of thinking affect accuracy of news detection (cognitive factor) but are also mediated by source assessment (background knowledge) and news frame detection (textual features).

Within the second approach, Tandoc and colleagues (2018b) have deduced from the respondents' answers that, similarly to conceptualization by Porshnev and colleagues, people base their judgment on "the self, the source, <and> the message" (p. 2754) on the first stage of judgment; they also use external authentication (combining the interpersonal/institutional sources with incidental/intentional motivation) when the first-step judgment leaves them hesitant (p. 2755–2766). However, to our regret, the data in the paper does not allow for explaining on what markers are used by the Singapore residents surveyed to detect the fakes; instead, users, the paper states, judge on the overall logic and tone of news pieces (p. 2754), without looking at sources cited or metatext. As we stated above, we see that exploring the role of, for instance, metatextual elements in audience judgment would surely add to better fact-checking practices.

Tandoc and colleagues' work demonstrate the striking lack of users' media educatedness in terms of truth construction elements in the text that would allow them to judge better and quicker on the true/false nature of news pieces, as well as their latent demand of clear-cut markers that would allow them to move from the whole-text level of "tone" or "logic" to peculiar signs of text authenticity. Moreover, Pennycook and Rand (2021) provenly insist that "poor truth discernment is associated with lack of careful reasoning and relevant knowledge, and the use of heuristics such as familiarity", not with political alignment (p. 388). To this, Albright (2017) adds one more aspect: Facts lose their engaging power, as "fact-based evidence is not relevant to a growing segment of the populace" (p. 87), which demands additional efforts from journalists – who, then, are forced to color their reporting with interpretations, at best.

Audience's diminishing engagement with the news vs. the growing need to distinguish between verifiable information and misleading one has, of course, much contributed to academic discussion about media consumers' news literacy. Tully and colleagues (2022) showed that, although participants were not particularly familiar with the concept of media literacy, some engaged in news literacy behaviors when confronted with news and information they perceived as unreliable, but only under certain circumstances. An important role of news literacy in fighting fakes was underlined by Morris and Yeoman (2023) by closely looking at how journalism educators perceived and interpreted news literacy. They, as well as Vartanova and Lukina (2022), argued that more professional training for teachers covering concepts related to news literacy and digital literacy, including misinformation vs. disinformation, are needed today, especially under the circumstances of the infodemic. However, a recent study showed that media literate individuals can use their knowledge and skills to create and circulate false information (Tully, 2022). Media literacy, therefore, should be seen as a complex area, with various possible audience effects and outcomes, and not as an indisputably positive phenomenon only.

All in all, there is a long-standing need for operational definitions and textual markers for various information disorders that would allow for countering fakes without critically hampering opinions, thus preserving the idea of the "blurred border" between falsehood and interpretation. A key question arises: Does the area of computational fake detection provide for them, being basically dedicated to finding such markers?

In the next section, we will discuss opportunities provided by computational methods and the advantages they present for fake detection and fake debunking in online environment, as well as challenges and limitations associated with computational fake detection.

3. Truly Useful Computations: A Demand for Non-Binary Logic

Computational communication studies of fake news are, in principle, a highly promising area of battling fakes, as they aim at creating successful automated filters of false information. Their fundamental contribution to fake news studies is exactly what "traditional" media studies of fakes lack – namely, finding the markers for fakes (predictors, features, proxies etc.) and making them work in practical news filtering. However, the computational research on fake detection also has some general shortcomings that can be diminished by closer collaboration with the "traditional" media research community.

First, and most important, computational communication studies of fake news nearly completely ignore the "blurred border" issue, mostly aiming at "yes/no" detection of fakes. It is determined by the nature of the RQs most often posed in such studies the way that they take into account neither international regulatory recommendations nor freedom of interpretation. Too often, computational linguistics sees the fake detection problem as a pure classification problem, in a mathematical way. This, unlike in the "classic" media and communication studies, nearly inevitably leads to binary "fake/non-fake" research designs, which hardly leaves room for the "blurred border" detection of multiple classes of distorted information. An example of the "dataset – preprocessing – features extraction – classifier training – content classification into false/truthful" is found in the review paper by Vishwakarma and Jain (2020).

Another example is the famous review of fake detection methods by Zhou and Zafarani (2018) that describes a highly diverse field of automated fake recognition. The authors have identified four major strategies for automation of fake detection, based on user traits, writing style, propagation style, and credibility markers. They scrupulously reconstruct the field that uses deception-based, style-based, language-structural, and attribute-based approaches – but all of them aiming at yes/no delineation of true text from false text, without the "blurred border" notion of truth gradations being taken into account. Moreover, even when researchers do extensive literature reviews on social, psychological, and technical factors of fake news spread, they end up in 1/0 classifications (as an example, see Shu et al., 2017).

Neither such studies follow conceptual differentiation of fake news as a genre from many other types of false information, from unintentional errors to rumors of no editorial origin to satire (Berghel, 2017; Mould, 2018). Even the key distinction between fact and interpretation stated as early as in 1921 by Charles Prestwich Scott of *The Guardian* as "comment is free but facts are sacred" (Scott, 1921) is disregarded. Even if the ultimate goal of detecting fakes is, indeed, distinguishing true news from fakes in a valid "yes/no" manner, research designs in this study area may be re-oriented to more nuanced classification of news (say, "true hard news", "soft news: balanced interpretation", "soft news: non-balanced interpretation", "fake news"). However, even hard news today does not always follow the inverted pyramid style, and the distinction between news and interpretation looks more and more problematic, demanding for further critical studies, at least for the sake of fake news detection. This lack of clarity is mirrored in fakes detection, as soon as there are more gradations than "yes/no." Thus, Lim (2018) found that fact-checkers agree well upon obvious truths/falsehoods but that the agreement rate is lower for "half true" or "mostly false" categories. Adding assessments based upon, e.g., one-/many-sided interpretations or, at least, fact/interpretation to the computational research designs may help eliminate such complications in automated fake detection, as well as inform manual fact-checking.

We would also like to underline that the conflict between the binary thinking on true/false fake detection and more complex epistemological considerations on the "blurred border" between fake and interpretation may be deeper than just a methodological issue emanating from the nature of each of the research areas. On one hand, there is a natural wish (and, thus, a public pressure) for simplicity and easiness of fake detection, for which the binary logic would be optimal; but, on the other hand, the centuries of struggles for media professionalism have created a spectrum of fact/interpretation modes in various genres of journalism, which undermines any such binarity. It is this fundamental opposition that is mirrored in the epistemological differences between media studies and computational battling of fakes.

Public pressure for clear-cut and simple detection of fake news also encounters the non-perfect state of automated methodologies which, *i.a.*, provide for potential "false positive" marking of real news as fake. We argue that each news piece has a right for presumption of innocence – that is, there is always a percentage of false positives in computational textual analysis that next-to-never reaches 100% accuracy. When dealt with news within big datasets, high enough error levels are seen as acceptable by the scholarly community who establish accuracy and precision metrics for accurate detection of the target data in computational linguistics, as well as accept "baseline" and "state-of-the-art" levels of reaching their maximums, which inevitably leaves space for errors. However, for a particular news piece of public importance, such an approach would be future-shaping, and the public would now wish to accept, say, a 20% probability of mistake, including taking real news for fake based on some rigid or formal parameters. This is why approaches aiming at flexibility in judging between falsehood and interpretation may also help avoid throwing out the baby with bathwater.

Second, not that long ago, even very careful and comprehensive review papers could not find "any attempts to clarify and delineate the concept of fake news in the information systems (IS) literature" (Khan et al., 2022: 3). The binary logic is sometimes accompanied by seeing fake news as a self-explanatory concept not needing further conceptualization, just as for media audiences. Khan and colleagues (2022) who, *i.a.*, rose this criticism have provided their own conceptualization of fakes as "a subset of false messages" (p. 4), distant enough, though, from the "blurred border" idea. Rare attempts of fake news typologies created in computational communication research (see, e.g., Rubin et al., 2015) would be dismissed by media scholars as done on flawed grounds.

Third, computational research focuses mostly on datasets compiled from social media data, while fakes, as stated above, mimic journalism at least as often as they mimic user-generated content. Moreover, Al-Rawi (2019) has shown that social media references to fakes mostly discuss biases in legacy media. And, as stated above, it is not social media texts primarily but journalism and media regulation that can help conceptualize the meaningful distortions in social media texts. The new large collaborations, like the one with Meta (Facebook/Instagram[1]) (Pasquetto et al., 2020), still aim at social media data; we insist that, in such projects, journalism-like fakes need to be separately addressed. This, in its turn, would imply the genre and metatext diversity discussed above as necessary to be taken into consideration.

Fourth, a major challenge in fighting fakes is taking into account wider social context behind mediated public discussions. Earlier, we had elaborated on significant underestimation of contextual knowledge in big social data studies, including computational communication research (see Bodrunova, 2020), showing that discussion context plays a mediating role in such research designs, shaping both the linkage between theories and research questions, on one hand, and interpretation of results, on the other. In fake news detection, though, contextual knowledge may play two more roles: It may help re-define fakes as such and provide computational scholars with contextualized understanding on the "blurred border" in delineating truth/interpretation from falsehood.

Fifth, the limitations of computational fake detection also arise from the nature of the units of analysis used in automated fact-checking. Thus, Caswell (2019) shows that "automated fact-checking… uses structured data (perhaps as annotation of training data) to represent features identifying a potentially newsworthy story or a fact claim, and each of which is therefore constrained by that data" (Caswell, 2019: 12). Thus, it is crucial to discuss the limits of big-data representations of facts and the implicit biases that arise from semantic granularity of "units" or "atoms" in "structured" journalistic texts, not only the explicit biases that arise from computational research designs (Caswell, 2019). Moreover, computational communication scholars are well aware of how many additional assumptions their research designs usually include, from proxies to excluded small-number data segments to too contextual (or way-too-decontextualized) interpretations of results (boyd & Crawford, 2012). This, *i.a.*, makes even the best works on computational fake detection hardly comparable and generalizable, including in how they conceptualize fakes.

Sixth, most often, computational studies of fakes remain just studies. Beyond the platforms that create active and working but publicly non-transparent filtering mechanisms, automated instruments elaborated by research groups do not find practical application, however precise they may be. It may happen that the detectors that may lead to delineating a spectrum of true/false texts would be even more complicated to construct and apply to practice; however, it may also be otherwise, if several steps of rough filtering are used to delineate fact-/falsehood-oriented texts from opinion pieces.

Thus, computational communication studies of fake news cannot so far provide for any widely shared typology of information disorders that would come "bottom up" from textual markers and correspond to the fact/interpretation and truth/falsehood dimensions, even if they have created many reliable methodological approaches on detecting individual fakes in a yes/no manner. Neither they efficiently help narrow down the growing gap between the rational and the mythological public mind, including the popular groups that rely on fakes and conspiracies in interpretation of reality (Bodrunova & Nepiyushchikh, 2019). They can be helpful when applied to real-world news streams, at least as "coarse filters", but they would better correspond to the reality of the "blurred border" diversity of truth construction means if they left space for journalistic interpretation in their models.

Today, there are several attempts to create taxonomies of news content that would allow for multi-level detection of fakes. Some of them (such as Molina et al., 2021) may provide interesting input for more sophisticated automated detection of fakes – and, thus, make the two research areas come closer to each other. However, even such multi-level taxonomies do not allow for the "mirror" multi-level operationalization of their categories for computational fake detection studies.

There have been attempts in the mid-2010s to detect the fakes based on spectra of fact/non-fact definitions. Most notably, a multi-class classification of the LIAR dataset (Wang, 2017) utilized categories such as true, mostly true, half-true, barely true, false, and pants-on-fire, a scheme adapted from PolitiFact's classification (true, mostly true, half true, mostly false, false, and pants on fire). The multi-class classification employed by BuzzFeed and BuzzFace Dataset (Silverman et al., 2016) used categories "mostly true", "mostly false", "mixture of true and false", and "no factual content." A multi-class classification system by BS Detector (https://github.com/selfagency/bs-detector) categorized information into Fake News, Satire, Extreme Bias, Conspiracy Theory, Rumor Mill, State News, Junk Science, Hate Group, Clickbait, and Proceed With Caution. Nakashole and Mitchell (2014) have made some efforts in this area of detecting fakes via regression models. One of the most impressive attempts in detecting fakes in social media content as based on multi-dimensional judgment is that by Mitra and Gilbert (2021) who curated the CREDBANK dataset of 60 million tweets and 96 days, grouped into 1,049 events, each event assigned a 30-dimensional vector of truthfulness labels and annotated by human coders as true/false using a 5-point Likert scale. However, the human-baseline coding for true/false is not guaranteed to come close to real truth, only to popularly perceived truth (remember how many people still think the sun goes around the Earth; Neuman, 2014). Categorizing on diverse grounds as in the BS Detector allows for teaching the machine to detect various types of distortions in presentation of reality but does not clearly delineate falsehoods from facts (e.g., true facts may be clickbait-written under well-known commercial pressures).

Moreover, all these studies are still arbitrary in how they judge true/false based on intuition by the detectors' creators; the "true/false" one-dimensional spectrum is nearly everywhere in these studies, except for, maybe, the BS Detector who tried to diversify the grounds for non-true detection with the regard of today's online phenomena such as clickbait or rumor mills, as well as ideologies within news production – from biases to state interventions. However, the grounds for detection of various types of suspicious information need to be properly systematized. Such attempts open floor for much more nuanced and multi-ground detection of various types of non-factual information; this is where less intuitive and more theoretical grounds for classification need to be introduced.

4. The Epistemic Framework for Detection of Fakes: Five pillars of Transparency and Trust-building

Delineating the gap between the "blurred border" logic in traditional media studies and the "yes/no" logic in computational communication ones, we propose an epistemological framework that would, to our viewpoint, help in diminishing this gap and, at the same time, allow for more nuanced detection of fakes without throwing the baby out with the bathwater. The goal of fakes detection needs to be seen as spectral/multi-class, and the approach needs to be applicable for the industrial detection of fakes, as well as helpful for media education of citizens.

To our viewpoint, the theoretical connection between the two research zones, as well as the practical detection of fakes and policymaking in this area, would benefit from thoroughly exploring five pillars of transparency and trust-building between the content creator and the reader. These pillars are:

1. True vs. false information;
2. Fact vs. interpretation;
3. Discrepancy vs. solidity;
4. Media texts vs. user-generated texts;
5. Human vs. AI/hybrid authorship.

Each pillar should provide for textual markers that are to be employed for detecting fakes. The pillars constitute the epistemic pipeline that may work in a step-by-step logic of fake detection.

True vs. false information: The proper use of the "yes/no" logic. The first stage of detection of fakes may truly well rely on detecting clear and proven factual falsehood (e.g., "the Amazon flows into the Pacific Ocean") from the rest of the analysed corpus. Here, one needs to be cautious in delineating facts/falsehoods from opinion statements, even if shared by large groups of scholars and publics (e.g. "Venezuela is an autocracy"). At this stage, not only deliberate falsehoods but also errors may be detected. We would call this stage a preliminary one, as, in its logic, it is close to pre-processing the data for errors and clear falsehoods. Other pillars, however, are way more complicated; but they need to be there to provide the epistemic dimensions for fake detection not only by machines but also by audiences themselves.

Fact vs. interpretation: The "blurred border" logic put in place. Of them, the fact vs. interpretation is the most complicated paradigm, which cannot be easily operationalized via text markers like information source or metadata (see Molina et al., 2021). A much more complicated spectrum that lies between what is seen by the media industry as fact and as pure interpretation and needs to be taken into account in all its complexity when detecting a "real" fact against all the rest. This is where the classes of more/less reliable texts need to be singled out in their relation to the extent to which interpretation is involved. This may imply the linguo-structural distinction between fact and various types of commentative text (opinion, statement, reflection etc.); but this would not be enough. The classifiers need to take into account at least two levels of markers beyond the lexical-grammatical appearances of fact vs. interpretation. On sentence and sub-sentence level, the classifiers need to be able to detect sarcasm, absurd, and rhetoric figures which may be taken by the machine for fake statements but become "false positives" in this case. On the level of whole text, genre structures and structural biases (including political ones) need to play a role, as they both for many decades have been influencing the construction of trust to media texts. Additionally, the divisions between disinformation, misinformation, and malinformation need to become seen as discursive, which will also provide for clearer textual markers. Thus, the classifiers need to work on several levels of the text composition, overlapping the markers step-by-step, in order to detect the classes to which a given text most probably belongs.

The nature of fake: A discrepancy test. Eliminating evident falsehoods and errors plus classifying the texts according their level of opinionation is still not enough, as none of these markers goes deep to the nature of a fake. The nature of a fake is fundamental discrepancy between the elements of reality, with two options for such a discrepancy – either between the parts of reality within a given text or a part of a text and a piece of reality beyond it. Thus, if a text can be tested against a body of knowledge available to large language models, it should be implemented – surely, with caution and human triple-checking. Moreover, the discrepancy test would look different for different classes of texts detected on the previous steps of the pipeline, as, say, hard news vs. news analysis would imply varying epistemic checks, while opinion columns would not need them at all – or very rarely.

Media texts vs. user-generated texts: Professional vs. witness metatextual evidence. If previous steps look complicated (but not unreachable), this step is easier to implement into the ML-based research but demands a slight conceptual shift in content creators' minds, as it potentially reshapes their relations with the audiences. The texts that contain harmful faked information may come either in the form of traditional media texts (and even mimic the existing mainstream media) or from non-media actors, from brands to bloggers to ordinary users. In the first case, metatext on the process of how a given news piece was created, what the editorial procedures were, and how the text relates to an editorial office would be extremely helpful in delineating fakes from well-produced news from the very beginning. Why not QR codes for editorial texts with deeper explanations on where the news piece came from and how? Even if QR codes would be too much, the metatextual markers of professionalism beyond the already-destroyed "inverted pyramid" style need to be seen as a key element of the struggle against false information. As to the texts generated beyond media entities, they have other types of metatext to employ, of which the signs of witness evidence may play the key role in proving the relations between reality and the information conveyed.

The human vs. AI authorship: Towards hybrid models of fact-checking. The final element of our epistemic pipeline is the figure of author. It is not only their linkage to a known or less known editorial office but also their profiles that create trust. Today, with the rapid development of AI authorship in media production, short news cease to bear authorship due to AI-assisted rewriting. This practice, however, creates open space for anonymized fake production – and, thus, should be eliminated. Moreover, recent studies have shown that hybrid (human+AI) models of stated authorship create more trust and are, indeed, more reliable than only human- or only AI-authored news (Opdahl et al., 2023): AI and human authors check each other, while the human author remains the responsibility bearer. Thus, revealing the authorship, linking it to a known media, brand, or personal profile, and complementing it with AI authorship may work as both a complex marker of true authorship and an additional discrepancy check.

If one looks closer, it would be easy to realize that the steps proposed above correspond more or less well to how a human would judge on the quality of a given news text. We first check it for evident errors and stupidities, then decide on whether it is news or opinion, then try to fundamentally judge on it against the body of knowledge we have and the genre of the text (as different genres require different logics of such judgment), and then doublecheck the source and metadata to make sure the news comes from a reliable source and is written according to standards and procedures we understand and accept as legitimate. The basic logic of human judgment needs to be brought back to communication studies in order to narrow the growing gap between the traditional and computational detection of fakes.

5. Conclusion and Recommendations

"Traditional" media and communication studies and computational communication research on fake news are both oriented to better detection of fakes – even if their propositions in reaching the rational ideal of news authenticity differ. The opposition between the socially-demanded simplicity of fake elimination that computational science tries to respond to and deeper epistemological considerations on what fakes are, as raised by media science, calls for a wider and inclusive discussion on a balanced approach to fake news detection.

For better results, the two research areas need to work together on the "blurred border" idea that seems to be shared but still not well enough operationalized by the media community. Computational studies may take the lead in revealing the markers for various classes of fakes, while these markers need to come primarily from normative journalistic text features, including professional metatext, and from the spectrum of possible fact/interpretation modes, not only from "true/false" statements judged linguistically or "against the commonsensical evidence."

For this, we have suggested above one (of many possible) way for further development. As it is clear from the UN statements on disinformation, "interpretations are free but false facts are eliminated", to rephrase C. P. Scott. Thus, put simply, fact and interpretation may be re-addressed as the concepts that allow, with proper operationalization, for distinguishing intentional disinformation from situational/unintentional misinforming, on one hand, and biased, one-sided, or even intentionally misleading but still factually unflawed reporting, on the other, as the most recent taxonomies created within more traditional media studies suggest. At the same time, one needs to look deeper into the nature of fake information and understand that there needs to be more pillars on which fake detection should fundamentally rest. We have proposed a five-pillar epistemological framework (that may convert into a five-step epistemic pipeline) for fake detection which implies the basic logic of human judgment upon the quality of a news text. It includes error elimination, separating fact-based from comment-based texts, a discrepancy test for solid correspondence to the body of extant knowledge, and authorship test, including metatextual professional/witness elements and proper markup of authorship, including the one by AI. If such complex models can be implemented, they would diminish the gap between the "blurred border" and "yes/no" logics of fake detection, as well as create systemic and non-arbitrary grounds for practical detection of fakes. However, even very sophisticated classifiers would not be enough if the professional community does not develop better metatextual markers for transparent news production.

For more "traditional" studies of fakes, elaboration of a sophisticated multi-class classification may be a way to operationalize the regulation-based "blurred border" understanding for further proof, elaboration, or criticism. For computational studies, it may provide input for better construction of multi-dimensional fake/non-fake classifiers oriented to many shades of fakes (especially if one adds the journalism/user-generated-content dimension which may lead to extracting different fact- and interpretation-based text features for them). For professional fact-checkers, it may provide an additional framework for proving their position. For them all, we hope, this may become a common ground for collaboration – again, just one of many possible.

For the audiences, in their turn, the democratic implications of more nuanced approaches to debunking fakes may ease the pressure of the demand for the less polluted news field, while preserving the freedom of speech. Automated filtering of fakes may efficiently complement manual fact-checking practices if (and, it seems, only if) it starts to correspond to multi-dimensional understandings of the fact/interpretation border stemming from freedom of interpretation that results from centuries of struggle for standards in public speech and journalism. Combined with more intense media education, especially in news production and how it may be recognized in verified media texts, orientation to delineating facts from opinions and falsehoods may lead to more rational, transparent, and calm news consumption.

Notes

[1] Meta Corp. platforms, including Facebook and Instagram, have been recognized as extremist and non-desired in Russia.

Funding Agency

This research is supported in full by Russian Science Foundation, project 25-18-00991 (2025 – 2027), year 1.

References

Al-Rawi, A. (2019). Gatekeeping Fake News Discourses on Mainstream Media Versus Social Media. *Social Science Computer*

Review, *37*(6), 687-704. https://doi.org/10.1177/0894439318795849

Albright, J. (2017). Welcome to the era of fake news. *Media and Communication*, *5*(2), 87-89. https://doi.org/10.17645/mac.v5i2.977

Baade, B. (2018). Fake News and International Law. *European Journal of International Law*, *29*(4), 1357-1376. https://doi.org/10.1093/ejil/chy071

Berghel, H. (2017). Lies, Damn Lies, and Fake News. *Computer*, *50*(2), 80-85. https://doi.org/10.1109/MC.2017.56

Bodrunova, S. S. (2020). The Boundaries of Context: Contextual Knowledge in Research on Networked Discussions. In A. Antonyuk & N. Basov (Eds.), *Networks in the Global World V* (pp. 165-179). Springer International Publishing. https://doi.org/10.1007/978-3-030-64877-0_11

Bodrunova, S. S., & Nepiyushchikh, D. (2019). Unhealthy communication on health: Discursive and ecosystemic features of opinion cumulation in the anti-vaccination discourse on Russian Telegram1. *World of Media*, (1), 75-108. https://doi.org/10.30547/worldofmedia.1.2024.4

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, *15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

Caswell, D. (2019). Structured Journalism and the Semantic Units of News. *Digital Journalism*, *7*(8), 1134-1156. https://doi.org/10.1080/21670811.2019.1651665

Graves, L. (2018). Boundaries Not Drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies*, *19*(5), 613-631. https://doi.org/10.1080/1461670X.2016.1196602

Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., González-Bailón, S., et al. (2019). Computational Communication Science: A Methodological Catalyzer for a Maturing Discipline. *International Journal of Communication (IJoC)*, *13*, 3912-3934. https://bit.ly/4nRfRyR

Humprecht, E. (2020). How Do They Debunk "Fake News"? A Cross-National Comparison of Transparency in Fact Checks. *Digital Journalism*, *8*(3), 310-327. https://doi.org/10.1080/21670811.2019.1691031

Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, *10*(1), 18-26. https://doi.org/10.1007/s13162-020-00161-0

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301-1326. https://doi.org/10.1177/1461444820959296

Khan, A., Brohman, K., & Addas, S. (2022). The anatomy of 'fake news': Studying false messages as digital objects. *Journal of Information Technology*, *37*(2), 122-143. https://doi.org/10.1177/02683962211037693

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., et al. (2018). The Science of Fake News. *Science*, *359*(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

Lien, C. H., Lee, J., & Tandoc Jr, E. C. (2022). Facing Fakes: Understanding Tech Platforms' Responses to Online Falsehoods. *Digital Journalism*, *10*(5), 761-780. https://doi.org/10.1080/21670811.2021.1982398

Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, *5*(3), 2053168018786848. https://doi.org/10.1177/2053168018786848

Mitra, T., & Gilbert, E. (2021). CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations. *Proceedings of the International AAAI Conference on Web and Social Media*, *9*(1), 258-267. https://doi.org/10.1609/icwsm.v9i1.14625

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, *65*(2), 180-212. https://doi.org/10.1177/0002764219878224

Morris, K., & Yeoman, F. (2023). Teaching Future Journalists the News: The Role of Journalism Educators in the News Literacy Movement. *Journalism Practice*, *17*(7), 1573-1590. https://doi.org/10.1080/17512786.2021.1992599

Mould, T. (2018). Introduction to the Special Issue on Fake News: Definitions and Approaches. *Journal of American Folklore*, *131*(522), 371-378. https://doi.org/10.5406/jamerfolk.131.522.0371

Nakashole, N., & Mitchell, T. (2014). Language-Aware Truth Assessment of Fact Candidates. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1009-1019). Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-1095

Neuman, S. (2014, February 14). 1 In 4 Americans thinks the Sun goes around the Earth, survey says. *NPR*. https://bit.ly/3WM7VnA

Nielsen, R. K., & Graves, L. (2017, October). "News you don't believe": Audience perspectives on fake news. *Reuters Institute for the Study of Journalism: Factsheet*. https://reutersinstitute.politics.ox.ac.uk/our-research/news-you-dont-believe-audience-perspectives-fake-news

Opdahl, A. L., Tessem, B., Dang-Nguyen, D.-T., Motta, E., Setty, V., Throndsen, E., Tverberg, A., et al. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, *146*, 102182. https://doi.org/10.1016/j.datak.2023.102182

Pasquetto, I. V., Swire-Thompson, B., Amazeen, M. A., Benevenuto, F., Brashier, N. M., Bond, R. M., Bozarth, L. C., et al. (2020). Tackling misinformation: What researchers could do with social media data. *Harvard Kennedy School Misinformation Review*, *1*(8), 1-14. https://bit.ly/49JZnFh

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388-402. https://doi.org/10.1016/j.tics.2021.02.007

Porshnev, A., Miltsov, A., Lokot, T., & Koltsova, O. (2021). Effects of Conspiracy Thinking Style, Framing and Political Interest on Accuracy of Fake News Recognition by Social Media Users: Evidence from Russia, Kazakhstan and Ukraine. In G. Meiselwitz (Ed.), *Social Computing and Social Media: Experience Design and Social Network Analysis* (pp. 341-357). Springer International Publishing. https://doi.org/10.1007/978-3-030-77626-8_23

Reese, S. D. (2023). Writing the Conceptual Article: A Practical Guide. *Digital Journalism*, *11*(7), 1195-1210. https://doi.org/10.1080/21670811.2021.2009353

Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4. https://doi.org/10.1002/pra2.2015.145052010083

Saldaña, M., & Vu, H. T. (2022). You Are Fake News! Factors Impacting Journalists' Debunking Behaviors on Social Media.

Digital Journalism, *10*(5), 823-842. https://doi.org/10.1080/21670811.2021.2004554

Schapals, A. K. (2018). Fake News. *Journalism Practice*, *12*(8), 976-985. https://doi.org/10.1080/17512786.2018.1511822

Scott, C. P. (1921, May 5). A hundred years. *The Manchester Guardian*. https://clck.ru/3QDZ49

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22-36. https://doi.org/10.1145/3137597.3137600

Silverman, C., Strapagiel, L., Shaban, H., Hall, E., & Singer-Vine, J. (2016, October 20). Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate. *Buzzfeed.News*. https://clck.ru/3QDYu7

Simons, G., & Manoilo, A. (2021). The what, how and why of fake news: An overview. *World of Media. Journal of Russian Media and Journalism Studies*, *2*, 35-55. https://doi.org/10.30547/worldofmedia.2.2021.2

Strasser, B. J. (2012). Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 85-87. https://doi.org/10.1016/j.shpsc.2011.10.009

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018a). Defining "Fake News". *Digital Journalism*, *6*(2), 137-153. https://doi.org/10.1080/21670811.2017.1360143

Tandoc Jr, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2018b). Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media & Society*, *20*(8), 2745-2763. https://doi.org/10.1177/1461444817731756

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467-473. https://doi.org/10.7326/M18-0850

Tully, M. (2022). Everyday News Use and Misinformation in Kenya. *Digital Journalism*, *10*(1), 109-127. https://doi.org/10.1080/21670811.2021.1912625

Tully, M., Madrid-Morales, D., Wasserman, H., Gondwe, G., & Ireri, K. (2022). Who is Responsible for Stopping the Spread of Misinformation? Examining Audience Perceptions of Responsibilities and Responses in Six Sub-Saharan African Countries. *Digital Journalism*, *10*(5), 679-697. https://doi.org/10.1080/21670811.2021.1965491

Tumber, H., & Waisbord, S. R. (2021). *The Routledge Companion to Media Disinformation and Populism*. Routledge. https://doi.org/10.4324/9781003004431

van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, *12*(2-3), 81-92. https://doi.org/10.1080/19312458.2018.1458084

Vartanova, E., & Lukina, M. (2022). The Triple Typology of Divide: Russia's Journalism Education in the Times of the COVID-19 Pandemic. *Journalism & Mass Communication Educator*, *77*(1), 74-91. https://doi.org/10.1177/10776958211053675

Vishwakarma, D. K., & Jain, C. (2020). Recent State-of-the-art of Fake News Detection: A Review. In *2020 international conference for emerging technology (INCET)* (pp. 1-6). IEEE. https://doi.org/10.1109/INCET49848.2020.9153985

Wagner, M. C., & Boczkowski, P. J. (2019). The Reception of Fake News: The Interpretations and Practices That Shape the Consumption of Perceived Misinformation. *Digital Journalism*, *7*(7), 870-885. https://doi.org/10.1080/21670811.2019.1653208

Waisbord, S. (2018). Truth is What Happens to News. *Journalism Studies*, *19*(13), 1866-1878. https://doi.org/10.1080/1461670X.2018.1492881

Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 422-426). Association for Computational Linguistics. https://doi.org/10.18653/v1/P17-2067

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking* (Council of Europe report DGI (2017) 09). Council of Europe. http://bit.ly/47w7Kmp

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025. https://doi.org/10.1016/j.ipm.2019.03.004

Zhou, X., & Zafarani, R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *arXiv preprint arXiv:1812.00315*. https://doi.org/10.48550/arXiv.1812.00315