

Assessing News Credibility: Misinformation Content Indicators

Paula Carvalho¹ · Danielle Caled¹ ·
Mário J. Silva¹ · Bruno Martins¹ · João
Paulo Carvalho¹ · Joaquim Carreira² ·
João Pedro Fonseca² · Teresa Gomes² ·
Pedro Camacho²

Received: date / Accepted: date

Abstract The development of explainable news credibility prediction models is critical both for fighting the viral propagation of misinformation and improving media literacy. This work investigates a variety of content indicators approaching different semantic and discourse dimensions, such as title representativeness, reasoning errors, and sentiment intensity. These indicators were inspired by a previous study conducted for English news, aimed at reaching a collective consensus on which indicators could be widely used for predicting news credibility. This new study, performed by a multi-disciplinary team, relies on a corpus of 80 news articles from Portuguese mainstream and alternative news media, which were annotated by junior and senior journalists. The assessment of the corpus annotations provides insight into the prevalence of different indicators in each type of news source. The results obtained for Portuguese correlate in most cases with the ones reported for English, which motivates the adoption of common standards for supporting the collaborative development of interoperable automatic misinformation detection approaches.

Keywords credibility indicators · content indicators · misinformation · disinformation · comparable news corpus · media literacy

1 Introduction

The negative societal impact of misinformation has attracted a growing interest from researchers, both from social and computer sciences, the latter particularly interested in identifying computational features and models to automatically detect false or misleading information presented as news, commonly called *fake news* [23]. Despite the promising outcomes of misinformation

P. Carvalho
E-mail: pcc@inesc-id.pt

¹INESC-ID Lisboa, Portugal · ²Lusa - Agência de Notícias de Portugal, Lisboa, Portugal

detection (MID), particularly the ones derived from deep learning techniques, most MID approaches do not directly contribute to improve media literacy, since the final consumer is not capable of understanding how a learning classifier arrives at a particular decision on the content credibility.

Rather than attempting to label or filter out a specific content as misleading, we aim to develop explainable models tailored to facilitate readers in analyzing, critically evaluating and ethically sharing information on the web, contributing both to the development of information literacy and fighting misinformation. To that end, it is crucial to firstly understand which credibility indicators could be identified in news content, and show association with misleading information (e.g. sentiment, title representativeness, sources of information, etc.).

In this paper, we describe a thorough annotation study that allowed to identify and assess a set of credibility indicators in Portuguese news texts, based on a corpus of 80 news articles, collected from both Portuguese mainstream and alternative media. Alternative news media seeks to address the gaps left by the mainstream news media, offering complementary, corrective and subjective perspectives on the topics or events reported in such news sources [10]. Since alternative media are not typically subject to any type of verification nor control, they are a potential source of misinformation [17].

The corpus annotation is based on a set of content indicators introduced by Zhang et al. [25] for assessing news articles' credibility, which were extensively discussed and tested in a collection of English news articles. Those indicators were then refined, based on a preliminary annotation study we conducted, which involved a team of media experts, linguists, and data scientists.

The annotation process was performed by three trained annotators with background in media studies, and an experienced journalist. The results provide an understanding of which characteristics are particularly productive and contribute the most for the perception of a news story as credible or misleading. We also investigate how the content indicators explored in our study interconnect with each other, and which ones are particularly informative for assessing credibility in news text.

To assess the reliability of the annotations in the corpus, we conducted an inter-annotator agreement (IAA) study, which helped understanding the consistency of the different indicators across annotators and how they relate with the expert annotations. Moreover, the IAA study allowed to evaluate the task complexity, and hence identifying the indicators that may require specific knowledge from annotators or access to information that is not provided in news text.

This work extends and consolidates the research previously carried out by Zhang et al. [25], namely by i) investigating new credibility indicators, ii) making adjustments to the content indicators that proved to be more difficult to recognize in the above-mentioned study, and by iii) applying the proposed methodology in a language other than English. Furthermore, our research demonstrates that the proposed indicators are a useful and powerful tool to

distinguish credible from misleading news, on the one hand, and mainstream from alternative news, on the other hand.

The remainder of the paper is structured as follows: Section 2 presents related work, Section 3 describes the data collection used in this study, and details the annotation process. Section 4 describes the results on the performed inter-annotator agreement study, provides some descriptive statistics on the annotated corpus, and presents the correlation results for a set of credibility indicators considered in our research. Finally, Section 5 highlights the main findings, and Section 6 discusses the research limitations and future research directions.

2 Related Work

Social media and alternative information sources allow users to produce a variety of news content, which is not subject to any type of verification nor control, often violating journalistic codes of ethics. These new sources of information are starters for real-time viralization of news stories, becoming a vehicle for the diffusion of misinformation to a global audience [23].

Most approaches for coping with misinformation employ journalists for fact-checking, aiming at assessing whether claims or statements are true. However, fact-checking is hard, expensive, and not feasible, given the amount of dynamic information disseminated through media channels [3]. Moreover, readers often access publishers directly instead of being mediated by third parties (e.g. fact-checking agencies). In addition, misinformation is often more viral and generally spreads faster than credible information, namely because of its engaging narrative [16, 20]. Furthermore, users are usually time-pressed news readers lacking media literacy skills, becoming “susceptible hosts” to misinformation [18].

Finding evidences that support the reader’s decision would be beneficial both for identifying potential biased or false information and preventing its further spreading. To this end, explanatory detection has become a trending research topic on misinformation detection [8]. Explainable machine learning models have been adopted to offer interpretable predictions on misinformation (e.g. Yang et al. 24), reporting why and how the systems arrived at a given prediction. In summary, models are not trained to infer the veracity of a news content, but to help readers make decisions on the content they are confronted with, by providing them with empirical evidence that may act as quality control indicators (e.g. Fuhr et al. 5).

Within this context, Fuhr et al. [5] proposed a model that adapt the concept of nutritional information to news content. The *information nutrition labels* would thus be analogous to the nutrition fact labels on food packages, allowing publishers and consumers to communicate and understand the underlying reasons that contribute to news credibility [25]. Among others, the label provides information on the following computable criteria: *factuality*, *virality*, *opinion*, *controversy*, *authority*, *technicality*, and *topicality* [5]. Aker et al. [2]

specifically addressed the sentiment indicator in news articles. They created a corpus composed of 250 fake and non-fake news articles with article-level sentiment indicators assigned by different paid annotators. The annotations performed suggest that fake articles are significantly more sentimental than the non-fake ones.

Following this idea, Kevin et al. [13] developed a browser plugin to assist online news readers in evaluating the quality of news content, based on the *source popularity*, *article popularity*, *ease of reading*, *sentiment*, *objectivity*, and *political bias*, considering both the news content and respective social network virality metadata. However, this study lacks an evaluation on how the indicators would help users to perceive misinformation. Besides, the authors do not provide any information about the relative relevance of the analyzed indicators, meaning that it is impossible to know which indicators would be more informative to the users.

Gollub, Potthast, and Stein [7] worked on the presentation of information quality indicators in an intuitive, unambiguous, and intelligible way. The authors designed textual metrics, organized into five categories or dimensions composing the information quality indicators. Then, a small crowdsourcing inquiry was conducted to assign interval ranges for each indicator. As an ideation study, no concrete implementation of the indicators was carried and, again, any analysis of its usefulness has been conducted.

To demonstrate that information quality indicators can also support automated fact-checking tasks, Agez et al. [1] proposed a model to handle the check-worthiness task (i.e. predicting which statements in a debate should be fact-checked). The authors considered a set of information attributes initially proposed by Fuhr et al. [5] as input features to their models. This work, rather than applying indicators as in the original conceptualization of the nutritional labels, employs them as explainable input features of a predictive model.

Information quality indicators were also used for inferring the credibility of news articles. In an exploratory study, Zhang et al. [25] defined and tested a set of context and content indicators, involving a team of media experts. Although the annotation study has involved a relatively small collection of news articles (40), it demonstrates that the proposed indicators lead to a critical reflection about the overall credibility of the reported events.

Though the current scientific efforts to define and implement new information quality indicators, the resources specifically tailored to deal with misinformation (namely, annotated *corpora*) are still scarce or even nonexistent in languages such as Portuguese. Moreover, it is difficult to assess which credibility indicators reported in literature are the most informative or efficient, since researchers typically adopt different methodological approaches and sets of indicators, making it difficult to perform a thorough comparative analysis.

3 Methods

This section describes the data collection, the annotation guidelines, the platform adopted for conducting the corpus annotation, and the profile of the annotators.

3.1 Data Collection

The corpus created for this study is composed of 80 news articles collected from Portuguese news sites: 15 mainstream, and 13 alternative news sources. The former includes authoritative news stories that follow journalistic codes of ethics, norms, and practices; the latter corresponds to news stories published in the blogosphere, where commitment to media standards is typically not required nor observed, being a potential source of misinformation.

Table 1 presents the news articles distribution across the sources crawled for this study. The collected news stories cover current hot topics at national and international levels (e.g. Coronavirus vaccine; 2020 United States presidential election) explored and disseminated by both mainstream and alternative media sources. When collecting data, we selected pairs of news that approach similar topics, and have an approximate length, in order to get a comparable corpus. Some studies state that credible news texts are significantly longer than non-credible (or fake) news [11], but we avoided using this as a primary differentiating feature.

The external information that might influence the annotators' assessments, including information about the publication source, and the website aesthetics [22], was removed. Then, the articles were randomly distributed into four subsets, comprising 20 documents each. Each annotator was assigned a subset of 20 different news articles, plus a subset of 20 news articles that were common to all the annotators. The latter corresponds to our golden collection, and will be used to assess the reliability of the annotations assigned to the entire corpus.

3.1.1 Corpus Characterization

Table 2 presents some statistics on the corpus, derived from a set of metrics often used in computational linguistics to obtain common textual features that would help characterizing the (mainstream and alternative) news texts included in our corpus. In particular, we considered simple quantitative metrics related to style and text complexity, such as the *title length* and *body text length*, which estimate the average number of words they comprise. We have also calculated the average number of sentences per document, and the average number of words per sentence.

Table 3 summarizes a set of linguistic metrics often used to measure *lexical richness*, specifically (i) *lexical diversity* (i.e. the ratio of different unique words to the total number of words) respecting both the title and the body content;

Mainstream news sources	#	Alternative news sources	#
Correio da Manhã	4	A Voz da Razão	4
Diário de Notícias	5	Bombeiros 24	2
Expresso	1	Desmascarar	4
Jornal de Negócios	1	Despertar Portugal	2
Jornal de Notícias	1	Direita Política	6
Jornal Económico	3	Estrategizando	1
Observador	8	Informação Incorreta	4
Público	2	Notícias Viriato	3
RTP Notícias	3	O Insurgente	1
Sábado	4	O Lado Oculto	6
Sic Notícias	2	Resistir	3
Sol	1	Semanário Extra	1
TSF	1	Tuga Press	4
TVI24	1		
Visão	2		
Total	39		41

Table 1 Distribution of the news articles composing our corpus.

	Mainstream		Alternative	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>
Title length (words)	11.1	3.0	11.2	5.3
Body length (words)	607.6	564.6	614.8	493.5
Sentences per document	22.4	28.6	24.4	17.8
Words per sentence	29.6	6.0	24.4	6.2

Table 2 Characterization of the corpus in terms of style and complexity.

(ii) *content diversity* (i.e. the ratio of different content words to the total number of content words), and (iii) *lexical redundancy* (i.e. the ratio of the total number of function or grammatical words to the total words (Zhou et al. 27)). We also include attributes specifically related to text credibility [28], namely the prominence of *modifiers* in text (i.e. the ratio of adjectives and adverbs to the total number of content words), the *lexical expressivity* (i.e. the ratio of adjectives and adverbs to the content words they usually modify, respectively nouns and verbs), and *pausality* (i.e. the number of punctuation marks over the total number of sentences).

Finally, Table 4 presents information on sentiment, which has been verified as a powerful feature for detecting misleading news [26]. In particular, we calculated the sentiment distribution (i.e. the ratio of potential sentiment words to the total number of words) and sentiment polarity distribution (i.e. the ratio of positive and negative words to the total number of sentiment words in text).

To get some of these statistics, all documents were initially tagged with *PoS* (*Part-of-Speech*) and sentiment information, through the application of the *CitiusTagger* [6], and *Sentilex* [19] and *OpLexicon* [21], respectively.

	Mainstream		Alternative	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>
Adjectives ratio	0.057	0.016	0.066	0.020
Adverbs ratio	0.047	0.011	0.053	0.017
Conjunction ratio	0.050	0.012	0.051	0.015
Nouns ratio	0.291	0.028	0.279	0.028
Verbs ratio	0.147	0.028	0.144	0.027
Title lexical diversity	0.751	0.097	0.736	0.057
Body lexical diversity	0.294	0.070	0.372	0.104
Content diversity	0.603	0.091	0.669	0.092
Redundancy	12.822	2.538	10.370	2.681
Expressivity	0.240	0.045	0.282	0.060
Modifiers	0.192	0.029	0.218	0.037
Pausality	4.102	1.567	2.980	0.937
Exclamation and question marks	0.004	0.007	0.039	0.066

Table 3 Linguistic characterization of the corpus.

	Mainstream		Alternative	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>
Title sentiment ratio	0.206	0.146	0.325	0.201
Title positive ratio	0.168	0.316	0.179	0.266
Title negative ratio	0.390	0.376	0.455	0.327
Body sentiment ratio	0.264	0.051	0.290	0.047
Body positive ratio	0.305	0.064	0.296	0.062
Body negative ratio	0.264	0.078	0.292	0.073

Table 4 Characterization of the corpus in terms of sentiment.

The results presented in Table 2 show that the mainstream and alternative news included in our corpus are similar in terms of document length and sentence complexity.

In what concerns lexical diversity, alternative articles make use of more different words in the news body (Table 3). On the other hand, mainstream texts use more grammatical words (*redundancy*) and punctuation signals (*pausality*). Particularly concerning punctuation, it is important to stress that non-declarative sentences ending with a question mark or an exclamation point are more common in alternative than in mainstream news texts. In fact, both signals are commonly identified in literature as features for assessing news credibility [9].

In what concerns *PoS* distribution, adjectives and adverbs are slightly more frequent in alternative news, which is also reflected by the *expressivity* and *modifiers ratio*. Sentiment words are also more usual in alternative news, particularly in news titles (Table 4).

Although the statistics previously presented show some differences between both types of news texts included in our corpus, no valid and supported conclusion can be drawn from these data. In fact, our sample is too small, and the differences observed in most cases are statistically insignificant. This re-

inforces the importance of pursuing a more in-depth study based on content indicators.

3.2 Annotation Guidelines

The annotation guidelines that we established for this study were inspired in the ones defined by Zhang et al. [25], who propose a set of indicators for assessing articles credibility, refined by a team of media experts. Those indicators explore both text content and external sources or metadata information. In our study, we only consider content indicators, since we are particularly interested in investigating how the information in text can signal (lack of) credibility and understanding to what extent those indicators are language or topic dependent.

Firstly, all content indicators described in the above-mentioned study were extensively discussed by our team of media experts, linguists, and data scientists, and assessed in a collection of 10 misleading news stories. This pilot annotation study allowed us to refine the guidelines, by (i) removing specific categories and subcategories (e.g. *scientific inference*) that do not seem relevant in our collection, (ii) adding other categories and subcategories that were not previously included (e.g. quotes or citations from *protagonists or witnesses*; some fallacies, like *personal attack*), (iii) distinguishing specific categories potentially pertinent in the context of our project (e.g. distinction between *acknowledged* - or intentional - and *undisclosed* - or unintentional - omission of the sources of information), and finally (iv) merging categories, such as quotes or citations from *outside experts* and *organizations*.

The final annotation guidelines describe two different, but complementary, annotation tasks, namely (i) assignment of semantic categories to a text segment, and (ii) replying to a set of close-ended questions about the overall perception of the article’s credibility. Henceforth, the former is mentioned as *fine-grained indicators*, and the latter as *coarse-indicators*.

3.2.1 Fine-grained indicators

In total, the code system includes 44 *fine-grained indicators*, organized into five different semantic categories, which were then classified into several sub-categories, as described in Table 5.

Title. The main purpose of news titles and headlines is to draw the reader’s attention to the story they briefly present. However, some titles can be misleading and unrepresentative of the body text, which may point to the lack of credibility of the news story being presented. Following [25], annotators were asked to identify the cases where the title is: *off-topic* (i.e. it explores a topic different than the one developed in the body); has a *different emphasis* (i.e. emphasizes information that does not correspond to the main issue approached in the body or takes a different stance than the body); carries *little*

Fine-grained indicators	
1 Title	2.1.3 Scientific study
1.1 Unrepresentative	2.1.4 Other
1.1.1 Off-topic	2.2 Secondary sources
1.1.2 Different emphasis	2.3 Omission of the sources
1.1.3 Little information	2.3.1 Declared omission
1.1.4 Different point of view	2.3.2 Undisclosed omission
1.1.5 Overstating claims	3 Fallacies
1.1.6 Understating claims	3.1 Appeal to action
1.1.7 Other	3.2 Appeal to authority
1.2 Clickbait strategies	3.3 Appeal to emotion
1.2.1 Listicle	3.4 Appeal to ignorance
1.2.2 Suspense	3.5 Appeal to fear
1.2.3 Provoking emotions	3.6 Appeal to nature
1.2.4 Hidden secret or trick	3.7 Personal attack
1.2.5 Challenges to the ego	3.8 Straw man
1.2.6 Cliffhanger	3.9 False dilemma
1.2.7 Inducing fear	3.10 Slippery slope
1.2.8 Other	3.11 Hasty generalization
2 Sources	4 Tone
2.1 Primary sources	4.1 Emotional charge
2.1.1 Expert/Organization	4.2 Exaggerated claims
2.1.2 Protagonist/Witness	5 Confidence
	5.1 Acknowledgement of uncertainty

Table 5 List of the fine-grained indicators considered in our annotation study.

information on the body; presents a *different point of view*; and *overstates* or *understates claims* presented in the body. Annotators were also asked to identify the potential *clickbait* tactics or strategies used to tempt the reader to click on the link to the story, in particular the ones commonly found in viral journalism (cf. Table 5).

Information Sources. The inclusion of citations is vital to ethical reporting, and to support the statements, claims and conclusions presented in news stories. In addition, citation of news sources provides accurate information about what happened or what was said, contributing for the story’s reliability [12, 15]. Therefore, annotators were asked to identify both primary and secondary sources, which can be used to investigate news’ credibility. The former includes *citation of scientific studies and reports*, *quotes from outside experts or organizations*, and *quotes from protagonists or witnesses* (i.e. individuals that are at the center of the event or issue, or have a personal knowledge of something and are able to testify to it.). On the other hand, secondary sources typically provide second-hand information, including, among others, cited news articles, reviews and news comments. Annotators were finally asked to identify (i) the cases where it is acknowledged that the information presented requires a quote or citation of a specific source to attest its reliability and accuracy, but it is omitted intentionally by the author (*declared omission of the information sources*), and (ii) the cases where this acknowledgement does not exist (*undisclosed omission of the information sources*).

Fallacies. Fallacies are reasoning errors that can be used deliberately or unintentionally by authors in argumentation to convince the readers of the validity of their arguments. In addition to the fallacies considered in the reference study (namely, *appeal to fear*, *appeal to nature*, *straw man*, *false dilemma*, and *slippery slope*), we have also taken into account *appeal to action*, *appeal to authority*, *appeal to emotion*, *appeal to ignorance*, *personal attack*, and *hasty generalization* fallacies, which proved to be relevant in the corpus explored in our pilot study.

Tone. Unlike the objective tone sought by most mainstream journalists, creators of alternative news typically use more subjectivity, seeking to strike an emotional connection with readers, in order to get an emotional reaction from them. Hence, annotators were asked to identify in text segments where authors employ a strong emotional tone or subjectivity (*emotionally charged tone*) and overstate or exaggerate their claims or points of view (*exaggerated claims*).

Acknowledgement of Uncertainty. The author acknowledges the possibility that the information presented could be unreliable somehow or analyzed from a different perspective.

3.2.2 Coarse-indicators

Additionally, annotators were asked to answer a set of 10 closed-ended questions (Table 6), to assess their overall perception of a set of properties related to news text credibility, which correspond to the *coarse-indicators*. These include dichotomous questions (*Yes/No*) and (five-point) *Likert*-scale questions. In addition to the questions 1 - 6, adopted from Zhang et al. [25], we added questions 7 and 8 to assess the overall sentiment and sentiment intensity conveyed in text, and questions 9 and 10 to evaluate the overall compliance with the linguistic and journalistic rules and standards, respectively.

3.3 Corpus Annotation

Three paid annotators were recruited for this task. All the annotators graduated either in Communication Sciences or in Media Studies, and two of them are currently attending a master program, one in Journalism and the other in News Media Management. Furthermore, all annotators completed at least one professional internship in a Portuguese leading news media company. The average age of the annotators is 23.6, and two annotators are male.

In addition to the annotations provided by the recruited annotators, the corpus was also annotated by an experienced journalist. This enabled evaluating how the recruited annotators' judgements agreed with the expert's judgements, and thus assessing the reliability of the annotated corpus. The expert has been a journalist for over 32 years. He was editor of several newspapers' sections (e.g. Society, Science, Politics) in a Portuguese leading daily newspaper and is currently editor-in-chief in the Portuguese News Agency (Lusa).

Coarse-indicator	Close-ended questions	Type
Overall credibility	How credible is the news text?	1-5
Title representativeness	Does the title accurately reflect the content of news text?	1-5
Clickbait title	How clickbaity is the title?	1-5
Single study	Does the news text focus on a particular scientific study?	Y/N
Supported conclusions	Are the claims, conclusions or points of view presented in the news text justifiable?	1-5
Convincing evidences	How convincing are the evidences presented in the news text?	1-5
Sentiment and opinions	Does the news text convey any sentiment or opinion?	Y/N
Sentiment intensity	How intense are the sentiment and emotions conveyed in news text?	1-5
Linguistic accuracy	How accurate is the news text from the linguistic point of view?	1-5
Journalistic standards	How rigorous is the news text regarding the journalistic rules and standards?	1-5

Table 6 Coarse-indicators and corresponding close-ended questions underlying them. The *Type* column shows the answers’ format and scale.

The corpus annotation was performed using *MAXQDA*, whose interface is illustrated in Figure 1¹. A training session, involving all annotators was conducted by an expert, who introduced them to the software, and demonstrated its main functionalities, namely the ones related to text codification (i.e. assignment of one or more *fine-grained indicators* to a selected text segment) and answering a list of close-ended questions related to the news article’s credibility (i.e. text overall assessment based on the *coarse-indicators* previously described). In *MAXQDA*, the *fine-grained indicators* correspond to *codes*, and the *coarse-indicators* are handled as *variables*.

To perform the previously described annotation task, the recruited annotators spent on average 15 minutes per news, while the expert dedicated about 1 hour to the analysis of each news article. This time difference is reflected on the expert’s meticulous work, who provided justifications for almost every decision made, and proposed some additional features to be explored in future work.

Most annotators reported that they consulted the guidelines during the annotation process more than five times. It must be stressed that guidelines include at least one illustrative example for each described code.

4 Corpus Analysis

This section describes the results on the inter-annotator agreement study performed, provides some descriptive statistics on the annotated corpus, and fi-

¹ VERBI Software. (2019). *MAXQDA 2020* [computer software]. Berlin, Germany: VERBI Software. Available at <https://www.maxqda.com/>.

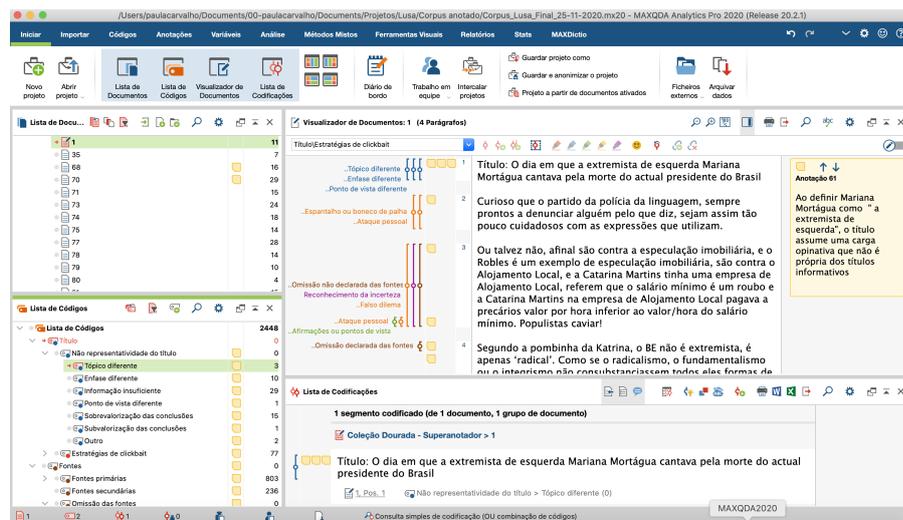


Fig. 1 MAXQDA interface for text annotation.

nally presents the correlation results for a set of credibility indicators considered in our research.

4.1 Inter-Annotator Agreement

To assess the reliability of the annotations in the corpus, we measured inter-annotator agreement (IAA), using Krippendorff's alpha, to compare the results we obtained with the ones reported in related studies using similar indicators [25]. In addition, the Krippendorff's alpha coefficient is suitable for comparing different types of data annotations (e.g. nominal, ordinal), including missing data, from any number of coders [14]. IAA relies on a data sample composed of 20 mainstream and alternative news articles, which were manually annotated by both the recruited annotators and the experienced journalist (*expert*). To measure to what degree the judgments produced by the annotators agree with the expert's assessments, we aggregated the judgments of the annotators as follows: (i) for ordinal data, we calculated the average across annotators, and (ii) for nominal data, we used the value chosen by the majority of annotators.

Table 7 shows the agreement rate regarding the *coarse-indicators* described in Section 3.2.2, and Table 8 shows the annotators' agreement regarding the explicit identification in text of the diversity of *fine-grained indicators* described in Section 3.2.1. In both cases, we present (i) the agreement obtained among the three annotators involved in this task (IAA), and (ii) the averaging agreement between the recruited annotators and the expert.

The agreement is substantial (0.6 - 0.8) for the majority of *coarse-indicators* considered in this study. In fact, when asked to make an overall judgement of text, the annotators mostly agree in distinguishing, for example, credible from

Coarse-indicator	Data type	IAA-RA	IAA-Expert
Overall credibility	ordinal	0.703	0.754
Title representativeness	ordinal	0.688	0.606
Clickbait title	ordinal	0.601	0.503
Single study	nominal	0.743	0.695
Supported conclusions	ordinal	0.597	0.759
Convincing evidences	ordinal	0.678	0.794
Sentiment and opinions	nominal	0.653	0.707
Sentiment intensity	ordinal	0.738	0.545
Linguistic accuracy	ordinal	0.128	0.788
Journalistic standards	ordinal	0.787	0.877

Table 7 IAA for *coarse-indicators*. *IAA-RA* refers to the agreement among the three recruited annotators, while *IAA-Expert* refers to the agreement between the aggregated annotations from the recruited annotators and the annotations made by the expert.

misleading news (*overall credibility*), accurate from inaccurate news titles (*title’s representativeness*), convincing from unpersuasive evidences (*convincing evidences*), and subjective from objective content (*sentiment and opinions; sentiment intensity*). Moreover, and probably due to their identical background in communication studies, annotators demonstrate a high degree of agreement regarding the judgement on the compliance with the journalistic standards. However, the same is not observed regarding their overall perception on the linguistic accuracy of news articles, leading to an agreement surprisingly slight (< 0.2). This low agreement suggests that annotators have different levels of linguistic demand and reinforces the idea that construct’s dimensions (e.g. grammar, vocabulary, and discourse) should be considered in future studies.

When comparing the judgments from recruited annotators to the expert’s, we verify that, except for *clickbait strategies* and *sentiment intensity*, which reached a moderate agreement (0.5 - 0.6), the agreement on the remaining *coarse-indicators* goes from substantial (0.6 - 0.8) to almost perfect (> 0.8). Again, the highest agreement achieved concerns the overall perception on the compliance of journalistic standards. With regard to the linguistic accuracy, the agreement increases significantly when considering the average rate given by all the annotators.

Overall, IAA results regarding the identification of generic clues related to text credibility are quite promising. Nevertheless, when we asked annotators to employ the *fine-grained indicators*, by explicitly identifying the text excerpts to which they apply, the agreement decreases significantly (cf. Table 8). In this case, most indicators had a moderate agreement (0.4 - 0.6), such as the ones involving the identification of both primary and secondary sources of information. Nevertheless, annotators tend to disagree in what concerns the identification of text segments where the source was (un)intentionally omitted. Still regarding the sources of information, IAA suggests the annotators and the expert seem to interpret the concept of *protagonist* and/or *witness* differently in this context.

Fine-grained indicators		IAA-RA	IAA-Expert
Unrepresentative Title	Different emphasis	0.210	0.649
	Little information	0.126	0.301
	Overstating claims	0.142	-0.114
Clickbait strategies	Defying convention	0.056	-0.054
	Provoking emotions	0.181	-0.054
	Inducing fear	0.454	0.331
	Cliffhanger to a story	0.056	0.331
Sources	Expert/Organization	0.541	0.480
	Protagonist/Witness	0.473	-0.161
	Scientific study	0.590	0.618
	Secondary sources	0.590	0.409
	Declared omission	-0.093	-0.114
	Undisclosed omission	-0.281	0.188
Fallacies	Appeal to action	0.562	0.831
	Appeal to emotion	0.291	-0.114
	Appeal to ignorance	0.056	0.458
	Appeal to fear	0.634	0.581
	Personal attack	0.359	0.494
	Straw man	-0.024	-0.147
	False dilemma	0.453	-0.182
	Slippery slope	0.399	0.304
	Hasty generalization	0.162	0.304
Tone	Emotionally charged	0.424	0.220
	Exaggerated claims	0.294	0.357
Confidence	Acknowledgement of uncertainty	0.433	-0.219

Table 8 IAA for *fine-grained indicators*. *IAA-RA* refers to IAA among the three recruited annotators, while *IAA-Expert* refers to the agreement between the aggregated annotations from the recruited annotators and the annotations made by the expert. Only the *fine-grained indicators* recognized both by the expert and at least one annotator are reported.

Regarding the news title, although annotators substantially agree in distinguishing representative from unrepresentative titles (cf. Table 7), they often disagree in identifying the potential issues underlying the title’s representativeness (cf. Table 8). Yet, particularly regarding the *emphasis* indicator, a strong agreement is observed between annotators and the expert.

Concerning fallacies, the agreement achieved among annotators in our study is quite encouraging, given the task’s subjectivity and complexity. For example, *appeal to fear* reached an agreement of 0.634, against 0.314 in the reference study (cf. Table 9). In addition, the agreement on the recognition of *false dilemma* clearly surpasses the low agreement reported in Zhang et al. [25], yet this rate drops significantly if we compare the annotations assigned by annotators to the ones assigned by the expert (cf. Table 8)².

² This could be explained given the discrepancy of annotations provided by annotators compared to the ones provided by the expert, who has only assigned this *fine-grained indicator* to three segments in the entire golden collection. An identical situation is inversely observed, for example, for *acknowledgement of uncertainty*, whose negative agreement be-

Coarse and Fine-grained indicators	Zhang’s	Ours
Title representativeness	0.367	0.688
Clickbait title	0.581	0.601
Sources - Single study	0.877	0.743
Sources - Expert /Organization	0.673	0.541
Sources - Expert/ Organization	0.283	0.541
Sources - Scientific study	0.763	0.590
Fallacies - Appeal to fear	0.314	0.634
Fallacies - Slippery Slope	0.478	0.399
Fallacies - Straw man	-0.096	-0.024
Fallacies - False dilemma	0.102	0.453
Tone - Exaggerated claims	0.235	0.294
Tone - Emotionally charged	0.098	0.424
Confidence - Acknowledgement of uncertainty	0.534	0.433
Confidence - Supported claims	-0.093	0.597
Inference - Convincing evidences	0.540	0.678

Table 9 Comparison of the IAA reported by Zhang et al. [25] against ours. Only the content indicators considered in both studies are reported.

Regarding the news article’s tone, we had a fair to moderate agreement; again the results obtained are higher than the ones reported in Zhang’s, specially in what concerns the recognition of an emotionally charged tone (0.433 against 0.098).

Globally, the inter-annotator agreement results reported in our study and the ones reported by Zhang et al. [25] have a moderate positive relationship ($r = 0.564$, measured using the Pearson correlation coefficient). Although the data collections cannot be directly compared, the lowest agreement in both studies concerns the *straw man* fallacy, which reinforces the difficulty in clearly identifying this fallacy in news articles. On the contrary, the highest agreement achieved is related to the identification of a particular scientific study underlying the news text.

4.2 Statistics on the Annotated Corpus

This section describes some statistics on the entire annotated corpus, considering both *coarse-indicators*, and *fine-grained indicators*.

4.2.1 Coarse-indicators statistics

Table 10 shows that annotators are capable of distinguishing mainstream from alternative news, and, based on their assessments, the former is significantly more credible than the latter. In a five-point rating scale, the overall credibility is 4.45 for mainstream news articles and only 1.65 for alternative news articles. This is also supported by the dissimilar values assigned to the recognition of

tween annotators and the expert is related to the fact that the latter has identified much more cases in text.

Coarse-indicator	Mainstream			Alternative		
	<i>mean</i>	<i>std</i>	<i>med</i>	<i>mean</i>	<i>std</i>	<i>med</i>
Overall credibility	4.45	1.13	5.00	1.65	0.94	1.00
Title's representativeness	4.36	0.98	5.00	2.18	1.29	2.00
Clickbait title	1.67	1.16	1.00	3.26	1.63	3.50
Supported claims	4.36	1.20	5.00	1.73	0.92	1.00
Convincing evidences	4.42	1.02	5.00	1.97	1.07	2.00
Sentiment intensity	1.23	0.67	1.00	3.43	1.34	4.00
Linguistic accuracy	4.73	0.54	5.00	3.50	1.18	3.00
Journalistic standards	4.48	0.89	4.00	1.58	0.97	1.00

Table 10 News article's overall perception regarding the five-point Likert-scale *coarse-indicators*.

Content Indicator	Mainstream		Alternative	
	#	%	#	%
Unrepresentative Title	14	1.64	47	2.95
Clickbait Title	19	2.22	58	3.64
Primary sources	524	61.36	279	17.50
Secondary sources	135	15.81	101	6.34
Omission of the sources	127	14.87	359	22.52
Fallacies	22	2.58	457	28.67
Tone	12	1.41	272	17.06
Acknowledgement of uncertainty	1	0.12	21	1.32
Total	854	100.00	1.594	100.00

Table 11 Absolute and relative distribution of *fine-grained indicators* (main categories) in the annotated corpus, contrasting mainstream and alternative news articles.

supported claims and convincing evidences in text, which have reached more than 4 points in mainstream news, and less than 2 points in alternative news. The opposite is observed with regard to sentiment, which seems to be much more expressive in alternative news. Regarding the overall compliance with the linguistic and journalistic rules, we observe that, according to the annotators, these standards are highly met in mainstream news, and fairly or poorly met in alternative news. Finally, in what concerns specifically the title, annotators considered that it is highly accurate in mainstream news (average rate = 4.36), and inaccurate in alternative news (average rate = 2.18).

4.2.2 Fine-grained indicators statistics

Table 11 provides information on the overall distribution of the main categories assigned by annotators to news articles from mainstream and alternative media sources. Globally, annotators have assigned 854 codes to the mainstream news articles, and 1.594 codes to the alternative news articles.

As expected, the indicators potentially associated with the lack of credibility or misinformation were mostly assigned to alternative news articles. For example, most fallacies considered in this study were found in alternative news, particularly *slippery slope*, *appeal to fear* and *hasty generalization*. Sim-

ilarly, an emotionally charged tone was chiefly recognized in alternative news, where annotators have also recognized an expressive number of cases involving overstatement and exaggerated claims (*tone*). The representativeness of such *fine-grained indicators* in mainstream news is almost insignificant.

Concerning the mention in text to information sources, the opposite is observed. Primary and secondary news sources are particularly productive in mainstream news, reinforcing the news credibility, while the omission of the source of information is more significant in alternative news. Nevertheless, annotators identified a huge number of cases where the required mention to a specific source to attest the information provided is also omitted in mainstream news.

4.3 Relationship between credibility indicators

Table 12 describes how the variables of this study correlate. For the ordinal indicators, we estimated the correlation using the Spearman’s rank correlation coefficient (ρ). For computing the correlation between each ordinal indicator and the news *source type* (ST), since the source type can be seen as a dichotomous variable, we employed the point biserial correlation coefficient (r_{pb}) assuming two possible values for ST , namely *mainstream* ($ST = 1$) and *alternative* ($ST = 0$).

The first conclusion to be drawn from our data is that mainstream news articles seem more credible, given the strong correlation between *source type* and *overall credibility* ($r_{pb} = 0.805$). As observed, the news article’s *overall credibility* is strongly dependent on the presentation of *supported claims* ($\rho = 0.883$) and *convincing evidences* ($\rho = 0.874$), which seem critical in mainstream news. Moreover, these aspects are positively related with the compliance of *journalistic standards*.

Conversely, a strong negative relationship is observed between *journalistic standards* and *sentiment intensity* ($\rho = -0.816$), meaning that the more subjective and emotional is the news text the less credible it seems to be. Regarding the *linguistic accuracy*, the correlation results suggest that alternative news tend to be more inaccurate from the linguistic point of view.

In regard to the title, the correlation results show that *representative titles* are more common in mainstream news ($r_{pb} = 0.688$), and that *clickbaitiness* is a strategy mostly used in news that do not fit the *journalistic standards* ($\rho = 0.584$).

Table 13 presents the results on the potential relationship between article credibility and the information sources mentioned in text. Despite the weak correlation observed in our corpus, it is possible to note two different trends: *undisclosed omission* of the source is negatively associated with *overall news credibility* ($\rho = -0.312$), while the explicit quotation or citation of the source of information (particularly primary sources, such as quotations from an *expert* or *organization*), is positively associated with the article’s *overall credibility* ($\rho = 0.607$).

	ST	OC	TR	CT	SC	CE	SI	LA	JS
ST									
OC	0.805								
TR	0.688	0.650							
CT	-0.486	-0.568	-0.720						
SC	0.778	0.883	0.681	-0.565					
CE	0.761	0.874	0.704	-0.556	0.873				
SI	-0.715	-0.746	-0.697	0.563	-0.704	-0.698			
LA	0.549	0.618	0.527	-0.384*	0.594	0.706	-0.439		
JS	0.840	0.897	0.707	-0.584	0.843	0.853	-0.816	0.646	

Table 12 Correlation between the source type (ST) and *coarse-indicators*: overall credibility (OC), title’s representativeness (TR), clickbait title (CT), supported claims (SC), convincing evidences (CE), sentiment intensity (SI), linguistic accuracy (LA), and journalistic standards (JS). All the correlations have a $p < 0.001$, with exception of LA-CT*, whose $p < 0.01$.

	OC	EO	PW	Sci	Sec	DO	UO
OC							
EO	0.607*						
PW	0.048	-0.114					
Sci	0.156	0.196***	-0.095				
Sec	0.183	0.208***	0.205***	0.229***			
DO	-0.154	-0.167	-0.082	0.109	-0.153		
UO	-0.312**	-0.070	0.032	0.060	0.049	0.263**	

Table 13 Spearman correlation between the perception of the news article’s overall credibility (OC) and the mention to information sources: experts and organizations (EO), protagonists and witnesses (PW), scientific studies (Sci), secondary sources (Sec), declared omission of the information sources (DO), and undisclosed omission of the information sources (UO). * $p < 0.001$, ** $p < 0.01$, *** $p < 0.05$.

	OC	AF	PA	SM	SS	HG
OC						
AF	-0.400*					
PA	-0.399*	0.085				
SM	-0.391*	0.328**	0.182			
SS	-0.519*	0.286**	0.135	0.480*		
HG	-0.388*	0.126	0.330**	0.285**	0.393*	

Table 14 Spearman correlation between the perception of the news article’s overall credibility (OC) and the most frequent fallacies the in corpus: appeal to fear (AF), personal attack (PA), straw man (SM), slippery slope (SS) and hasty generalization (HG). * $p < 0.001$, ** $p < 0.01$.

Table 14 presents the correlation considering the news article’s overall credibility, and the most frequent fallacies in our annotated corpus. A negative relationship can be observed between all fallacies and the news article’s credibility. In our corpus, the higher correlation coefficient achieved applies to *slippery slope*, which has moderate to high negative correlation with the article’s *overall credibility* ($\rho = -0.519$).

4.4 Relevance of Credibility Indicators

We evaluated whether the *coarse-indicators* and the main categories associated with the *fine-grained indicators* could be used to predict the article’s credibility, through a backward step-wise linear regression, using the *overall credibility* as the dependent variable and the remaining indicators as independent variables. We carried out this analysis over the golden collection, by incrementally removing the variables that caused the most statistically insignificant deterioration on the model. Regarding the *coarse-indicators*, 4 variables remained: *title’s representativeness*, *clickbait title*, *sentiment intensity*, and *linguistic accuracy*. Regarding the *fine-grained indicators*, the resulting variables are: *fallacies* and *primary sources*. We obtained, for *coarse-indicators* and *fine-grained indicators*, respectively: $R^2 = 0.857$ and $R^2 = 0.504$; *Adjusted R^2* = 0.853 and *Adjusted R^2* = 0.491, after the model convergence; and *F-statistic* = 230.0 and *F-statistic* = 39.05 ($p < 0.001$).

5 Discussion and Main Conclusions

This annotation study was based on a focused set of already tested credibility indicators [25], which allowed assessing their pertinence and adequacy in a different collection of news articles from a different language, and refining some categories based on previous experiments. Having similar guidelines and assessing instruments made it possible to compare the results achieved in both studies, and to define the next research steps regarding the detection of misinformation in news articles in a solid manner.

Overall, our research shows that the guidelines adopted from Zhang et al. [25] apply to other news topics and languages, specifically Portuguese. It also demonstrates that both *coarse-* and *fine-grained indicators* allow distinguishing mainstream from alternative news. In fact, we found several indicators that are closely related (in terms of agreement) with the ones reported in the above-mentioned study, which reinforces the reliability of the annotated corpus, and provides thorough information on the task complexity. Moreover, the results suggest that the alternative news included in our corpus are untrustworthy, presenting a sort of characteristics that are not usually found in news collected from mainstream media.

The corpus annotation made it possible to determine which indicators are more frequent in both mainstream and alternative news, and content to understand which ones are easier to recognize, based on the inter-annotator agreement study we performed. Regarding frequency, the mention to the sources of information in news articles is mostly found in mainstream news, while alternative news sources typically omit them. In addition, the latter often makes use of an emotionally charged tone, and overstates the conclusions and claims presented in text, presenting frequently different sorts of fallacies.

IAA results show that annotators are capable of globally distinguishing credible from misleading news. However, when we asked them to identify in

text specific credibility indicators, the task becomes harder and more complex. In a five-point Likert scale (where 1 represents “no difficult” and 5 “extremely difficult”), most annotators selected 4 to characterize the task complexity, in an online anonymous survey they responded after completing the annotation process. Additionally, they mentioned that the *fine-grained indicators* most difficult to recognize in news articles were fallacies, in particular *appeal to nature*, *slippery slope* and *straw man*. This is corroborated by the low agreement achieved regarding the identification of some fallacies, with particular emphasis to *straw man*.

When taking into account the annotations provided by our expert, we also found specific cases that may be out of reach for non-expert annotators, or that require access to information that goes beyond text, such as the *acknowledgement of uncertainty*, some fallacies (e.g. *straw man* and *false dilemma*), and specific sources of information (e.g. *protagonists and witnesses*).

However, as mentioned before, the annotation results also show that some of these *fine-grained indicators*, namely the ones related with reasoning errors are prevalent in the misleading news articles, and could be used as features to assess text credibility. Hence, a further investment on a deeper analysis and formalization of such discourse strategies is needed.

According to the annotators’ feedback, the easiest *fine-grained indicators* to recognize in news texts are the *sources of information*, particularly *primary sources*. In fact, annotators are globally capable of identifying different types of mentions to explicit sources of information in text. However, in what concerns the cases where the source of information is required in text, but was omitted by the author, annotators tend to often disagree. Disagreement among annotators, on the one hand, and between annotators and the expert, on the other hand, reinforces the importance of discussing and clarifying such indicators in future research, particularly *undisclosed omission*, given its representativity in the corpus.

IAA results also show that annotators are highly sensitive to *journalistic standards*, which is certainly linked to their specific background. In further experiments, it would be important to extend the annotation study to common news readers, to ascertain if this feature is also perceptible by non-experts.

On the contrary, the annotators’ judgments on the *linguistic accuracy* of text are extremely divergent, despite the linguistic differences underlying mainstream and alternative news articles. This reinforces the importance of refining the linguistic assessment considering more specific dimensions (e.g. grammar, vocabulary, punctuation, etc.), namely because it appears to be a relevant indicator for misinformation detection, together with *title representativeness*, *title clickbaitiness* and *sentiment intensity*.

In line with the research results reported in literature [2], our study shows that sentiment plays an important role on the identification of credible news. In fact, the degree of expressiveness, sentiment and subjectivity in text (supported, among other evidences, by the usage of expressive punctuation, sentiment polarity and intensity) could be explored in future research to help dis-

tinguishing not only credible from misleading news, but also identifying news styles, namely the ones that make use of sensationalism in their reporting.

6 Limitations and Future Work

Our research shows that the content indicators developed by Zhang et al. [25] can be applicable to other languages, in particular Portuguese, answering to a potential limitation they reported in their study. Nevertheless, it should be stressed that the comparison results with the above-mentioned study are only indicative and approximate, namely because the data collections are different and the content indicators are not completely coincident.

Furthermore, the generalizability of the results is limited by the corpus size and the news sources considered. Since we were interested in identifying clues associated with misinformation, we intentionally selected mainstream news sites that typically accomplish the journalistic standards, and alternative news sites that do the opposite. Hence, although the alternative news in our corpus seem untrustworthy, this conclusion cannot be generalized to the totality of alternative news media. In fact, there is a variety of news and media outlets that are not covered in our study.

Further research involves the enlargement of the annotated corpus, focusing on the indicators that demonstrate to be the most relevant in the present research. Additional news sources will be automatically crawled from the Portuguese Web Archive³, and the annotation will involve the participation of common news readers, adopting eventually strategies like the ones reported in [4].

Our ultimate goal is to use the enlarged corpus to train an explainable model for news credibility prediction and later embed it in a new tool to be used for improving media literacy and assisting news consumers in coping with disinformation.

Acknowledgements

This work was supported by Portugal2020 (POCI-05-5762-FSE-000217), and by Fundação para a Ciência e a Tecnologia (PTDC/CCI-CIF/32607/2017, UIDB/50021/2020, and SFRH/BD/145561/2019).

Conflict of interest

The authors declare that they have no conflict of interest.

³ <https://arquivo.pt/?l=en>

References

1. Agez R, Bosc C, Lespagnol C, Petitcol N, Mothe J (2018) IRIT at Check-That! 2018. In: Cappellato L, Ferro N, Nie J, Soulier L (eds) Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, CEUR Workshop Proceedings, vol 2125
2. Aker A, Hamacher M, Nti A, Smets A, Gravenkamp H, Erdmann J, Mayer S, Serong J, Welpinghus A, Marchi F (2019) Corpus of news articles annotated with article level sentiment. In: NewsIR@ SIGIR, pp 30–35
3. Alaphilippe A, Gizikis A, Hanot C, Bontcheva K (2019) Automated tackling of disinformation. Tech. rep., European Parliamentary Research Service, DOI 10.2861/368879
4. Bhuiyan MM, Zhang AX, Sehat CM, Mitra T (2020) Investigating differences in crowdsourced news credibility assessment: Raters, tasks, and expert criteria. Proceedings ACM on Human-Computer Interaction 4(CSCW2), DOI 10.1145/3415164, URL <https://doi.org/10.1145/3415164>
5. Fuhr N, Giachanou A, Grefenstette G, Gurevych I, Hanselowski A, Jarvelin K, Jones R, Liu Y, Mothe J, Nejd W, et al (2018) An information nutritional label for online documents. SIGIR Forum 51(3):46–66, DOI 10.1145/3190580.3190588
6. Garcia M, Gamallo P (2015) Yet another suite of multilingual NLP tools. In: Sierra-Rodríguez JL, Leal JP, Simões A (eds) Languages, Applications and Technologies, Springer International Publishing, Cham, pp 65–75
7. Gollub T, Potthast M, Stein B (2018) Shaping the information nutrition label. In: Albakour D, Corney D, Gonzalo J, Martinez-Alvarez M, Poblete B, Valochas A (eds) Proceedings of the Second International Workshop on Recent Trends in News Information Retrieval co-located with 40th European Conference on Information Retrieval (ECIR 2018), CEUR-WS.org, Grenoble, France, CEUR Workshop Proceedings, vol 2079, pp 9–11
8. Guo B, Ding Y, Yao L, Liang Y, Yu Z (2020) The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys 53(4), DOI 10.1145/3393880, URL <https://doi.org/10.1145/3393880>
9. Hardalov M, Koychev I, Nakov P (2016) In search of credible news. In: Dichev C, Agre G (eds) Artificial Intelligence: Methodology, Systems, and Applications, Springer International Publishing, Cham, pp 172–180, DOI 10.1007/978-3-319-44748-3_17
10. Holt K, Figenschou TU, Frischlich L (2019) Key dimensions of alternative news media. Digital Journalism 7(7):860–869, DOI 10.1080/21670811.2019.1625715, URL <https://doi.org/10.1080/21670811.2019.1625715>
11. Horne BD, Adali S (2017) This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. arXiv preprint arXiv:170309398

12. Johnson KA, Wiedenbeck S (2009) Enhancing perceived credibility of citizen journalism web sites. *Journalism & Mass Communication Quarterly* 86(2):332–348, DOI 10.1177/107769900908600205, URL <https://doi.org/10.1177/107769900908600205>, <https://doi.org/10.1177/107769900908600205>
13. Kevin V, Högden B, Schwenger C, Şahan A, Madan N, Aggarwal P, Bangaru A, Muradov F, Aker A (2018) Information nutrition labels: A plugin for online news evaluation. In: *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, Association for Computational Linguistics, Brussels, Belgium, pp 28–33, DOI 10.18653/v1/W18-5505
14. Krippendorff K (2004) Measuring the reliability of qualitative text analysis data. *Quality and quantity* 38:787–800, DOI 10.1007/s11135-004-8107-7, URL <https://doi.org/10.1007/s11135-004-8107-7>
15. Lashmar P (2019) Sources and source relations. *The International Encyclopedia of Journalism Studies* pp 1–9, DOI <https://doi.org/10.1002/9781118841570.iejs0270>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118841570.iejs0270>, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118841570.iejs0270>
16. Marwick A, Lewis R (2017) Media manipulation and disinformation online. Tech. rep., Data & Society Research Institute, New York, NY, USA
17. Robertson CT, Mourão RR (2020) Faking alternative journalism? an analysis of self-presentations of “fake news” sites. *Digital Journalism* 8(8):1011–1029, DOI 10.1080/21670811.2020.1743193, URL <https://doi.org/10.1080/21670811.2020.1743193>, <https://doi.org/10.1080/21670811.2020.1743193>
18. Rubin VL (2019) Disinformation and misinformation triangle. *Journal of Documentation* 75:1013–1034, DOI 10.1108/JD-12-2018-0209, URL <https://doi.org/10.1108/JD-12-2018-0209>
19. Silva MJ, Carvalho P, Sarmiento L (2012) Building a sentiment lexicon for social judgement mining. In: Caseli H, Villavicencio A, Teixeira A, Perdigão F (eds) *Computational Processing of the Portuguese Language*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 218–228, DOI 10.1007/978-3-642-28885-2_25
20. Silverman C (2015) Lies, damn lies and viral content. Tech. rep., Tow Center for Digital Journalism, Columbia University, DOI 10.7916/D8Q81RHH
21. Souza M, Vieira R (2012) Sentiment analysis on twitter data for Portuguese language. In: *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*, Springer-Verlag, Berlin, Heidelberg, PROPOR’12, pp 241–247, DOI 10.1007/978-3-642-28885-2_28, URL http://dx.doi.org/10.1007/978-3-642-28885-2_28
22. Spillane B, Lawless S, Wade V (2020) The impact of increasing and decreasing the professionalism of news webpage aesthetics on the perception of bias in news articles. In: Kurosu M (ed) *Human-Computer Interaction. Design and User Experience*, Springer International Publishing, Cham, pp 686–710, DOI 10.1007/978-3-030-49059-1_50

23. Tandoc E, Lim ZW, Ling R (2018) Defining “fake news”. *Digital Journalism* 6(2):137–153, DOI 10.1080/21670811.2017.1360143, URL <https://doi.org/10.1080/21670811.2017.1360143>, <https://doi.org/10.1080/21670811.2017.1360143>
24. Yang F, Pentyala SK, Mohseni S, Du M, Yuan H, Linder R, Ragan ED, Ji S, Hu XB (2019) Xfake: Explainable fake news detector with visualizations. In: *The World Wide Web Conference*, Association for Computing Machinery, New York, NY, USA, WWW '19, p 3600–3604, DOI 10.1145/3308558.3314119, URL <https://doi.org/10.1145/3308558.3314119>
25. Zhang AX, Ranganathan A, Metz SE, Appling S, Sehat CM, Gilmore N, Adams NB, Vincent E, Lee J, Robbins M, Bice E, Hawke S, Karger D, Mina AX (2018) A structured response to misinformation: Defining and annotating credibility indicators in news articles. In: *Companion Proceedings of the The Web Conference 2018*, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, WWW '18, p 603–612, DOI 10.1145/3184558.3188731
26. 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 <http://www.sciencedirect.com/science/article/pii/S0306457318306794>
27. Zhou L, Burgoon JK, Twitchell DP, Qin T, Nunamaker Jr JF (2004) A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems* 20(4):139–166, DOI 10.1080/07421222.2004.11045779, URL <https://doi.org/10.1080/07421222.2004.11045779>, <https://doi.org/10.1080/07421222.2004.11045779>
28. Zhou X, Zafarani R (2020) A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys* 53(5), DOI 10.1145/3395046, URL <https://doi.org/10.1145/3395046>