Online News Verification: An AI-Based Platform for Assessing and Visualizing the Reliability of Online News Articles

Anne Grohnert, Simon Burkard, Michael John, Christian Giertz, Stefan Klose, Andreas Billig, Amelie Schmidt-Colberg and Maryam Abdullahi
Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany
fi

Keywords: Online News Verification, Web Information Analysis, Fake News Detection, Disinformation, Reliability.

Abstract: Assessing the reliability of online news articles poses a significant challenge for users. This paper presents a novel digital platform that enables users to analyze German-language news articles based on various reliability-related aspects, including opinion strength, sentiment, and article dissemination. Unlike many existing approaches focused solely on detecting fake news, this platform emphasizes the comparative analysis and visualization of relevant reliability indicators across articles from different publishers. The paper provides a comprehensive overview of the current state-of-art describing various existing approaches for the detection and presentation of disinformational online content before presenting the technical system architecture and user interfaces of the designed platform. A concluding user evaluation reveals some limitations and opportunities for further developments, but showed generally positive feedback on the platform's diverse analysis criteria and visual presentation to support users in assessing the credibility of news articles. Potential future applications range from evaluating article neutrality to verifying citations in academic contexts.

1 INTRODUCTION

Assessing the credibility and reliability of online content is often difficult for users. Valuable support in this task could be provided by an intelligent system that assesses the trustworthiness of online content and offers assistance in interpreting the characteristics of online articles with regard to their credibility. This applies not only to the assessment of user-generated posts on social media platforms but also to the analysis of the credibility of journalistic articles from reputable as well as less reputable news publishers. A tool for analyzing the reliability of online news articles could, for example, help in assessing the objectivity of online articles and the citability of journalistic sources, or assist in analyzing the dissemination paths of certain online content.

Existing research focuses mostly on the development of efficient methods and algorithms for the automated detection and classification of disinformative content (fake news), especially in the context of social media, but less on the development of systems that analyze and compare news articles with respect to various reliability features visualizing also the results of the aspectoriented analysis.

This paper presents a novel digital platform on which users can examine individual German-language news articles with respect to various aspects of disinformation to gain a structured, uniform, and personalized overview of news articles from different news publishers. The paper begins with a detailed overview of existing methods and the current state of research in the field of disinformation detection. Afterwards, the technical system architecture of the developed platform is presented and the components and interfaces involved are described, e.g., for AI-based assessment of article trustworthiness, determination of opinion strength, and visualization of article dissemination paths. A concluding summary of a user evaluation highlights potentials and opportunities for further development.

2 STATE OF THE ART

Nowadays exist several approaches to detect fake news. These are divided into manual checks, automated procedures and systems that combine several approaches.

Grohnert, A., Burkard, S., John, M., Giertz, C., Klose, S., Billig, A., Schmidt-Colberg, A. and Abdullahi, M.
Online News Verification: An AI-Based Platform for Assessing and Visualizing the Reliability of Online News Articles.
DOI: 10.5220/0013778100003985

In *Proceedings of the 21st International Conference on Web Information Systems and Technologies (WEBIST 2025)*, pages 412-419
ISBN: 978-989-758-772-6; ISSN: 2184-3252

2.1 Manual Disinformation Detection

Various fact checking organizations rely on manual human-based analysis of news content, such as Correctiv (CORRECTIV, 2025) or Faktenfuchs (BR24, 2025) in German-language media, PolitiFact (Poynter, 2025) as US fact-checking platform or Full Fact (FullFact, 2025) in the UK. They aim to identify, verify and correct misinformation and disinformation in the public sphere helping to shape public opinion by providing fact-based information. They verify claimed facts or political statements circulating in the media, social networks and publish explanatory articles why the provided information is false, misleading or correct. Many of these organizations work according to a transparent code of conduct (e.g. the International Fact-Checking Network - IFCN of Pointer (IFCN, 2025)). They disclose how they work, who funds them and on what basis judgments are made.

2.2 Approaches for Automatic Detection of Disinformation

Besides manual fact checking various (scientific) approaches exist at present for the automatic detection of disinformation. Table 1 shows the various approaches, the targeted detection focus and a few examples of state-of-the-art existing systems.

Fact Checking. In the field of automated fact-checking, statements are verified by comparing them with facts or knowledge bases. Typical methods include claim detection (identifying verifiable claims), claim matching (comparing with known facts, e.g., from PolitiFact, Snopes, Wikipedia, Wikidata) and natural language inference (NLI), assessing whether a text supports, contradicts or is neutral towards a claim. For example, dEFEND is an explainable fake news detection system that analyzes both the news content for particularly check-worthy sentences and processes user comments to filter out opinions and indications of false information (e.g. skepticism, references to fact checks) (Shu et al., 2019). By analyzing news content and comments together, relevant key sentences and comments are identified that serve as an explanation. As a result, the model not only provides a classification (fake vs. real), but also shows the specific sentences and comments that influenced its decision.

Source Credibility. Source analysis (source credibility) involves assessing the reliability of the original source (website, author, domain). Among other characteristics, the domain reputation (e.g. .edu vs. .xyz), but also the history of the source (e.g. existence of false statements in the past) are considered. Browser plugins such as NewsGuard (NewsGuard Technologies, 2025) or TrustyTweet (Hartwig and Reuter, 2019) can be used directly for source credibility analysis without having to leave the website. NewsGuard evaluate news sites according to nine publicly accessible criteria that are based on journalistic standards and source transparency (e.g. clear separation of opinion and news, regular publications and transparency about provider, publisher and ownership structure). TrustyTweet analyzes and evaluates Twitter/X posts for credibility through source verification and the use of AI. Warnings or trust ratings are displayed directly below the posts.

Style Analysis. The aim of style analysis (writing style, linguistic features) is to identify disinformation through characteristic linguistic patterns. Characteristics such as the use of excessive adjectives, superlatives and exclamation marks as well as low objectivity and high subjectivity or the use of sensationalistic language are taken into account. The FakeFinder app recognizes fake news from the Twitter livestream and warns users in real time. It is based on the small open-source NLP language model ALBERT from Google AI (Tian et al., 2020). Grover is an AI model for generating and recognizing fake news (Zellers et al., 2019). It was trained using a transformer model on a data set of real journalistic articles. Grover analysis the style, structure, metadata and linguistic features and classifies whether the text probably originates from a human journalist, from Grover itself or from another AI model.

Sentiment Analysis. Sentiment analysis and emotion recognition are about recognizing emotionally content that is typical for propaganda or disinformation. The dominant emotions here are fear, anger or outrage and the language used is often polarizing. DeepMoji is a neural language model that was developed to recognize emotions and moods in texts based on the emojis used (Felbo et al., 2017). Around 1 billion tweets on Twitter with emojis as labels served as training data and 64 of the most frequently used emojis served as target classes for the modeling.

Social Signals. In the area of social context analysis (social signals), the behavior of dissemination and interactions in social networks is examined. Models such as Propagation Tree Analysis, Graph Neural Networks (GNNs) or Temporal Pattern Modeling (e.g. LSTM, Transformers) are used. Botometer (formerly: BotOrNot) is an online tool designed to detect social media bots on Twitter/X (Yang et al., 2022). It evaluates how probable it is that an account is an automated Bot.

Hoaxy is a web platform and visualization tool that shows how fake news and fact checks spread in

Table 1: Different approaches for disinformation detection.

Approach	Detection Focus	Example System(s)
(1) Fact Checking	Veracity of Claims	dEFEND
(2) Source Credibility	Trustworthiness of the Source	NewsGuard, TrustyTweet
(3) Style Analysis	Linguistic Patterns	Grover, FakeFinder
(4) Sentiment Analysis	Emotional Manipulation	DeepMoji
(5) Social Signals	Dissemination Patterns	Botometer, Hoaxy
(6) Multimodal	Inconsistency between Image and Text	VisualBERT, VilBERT
(7) Knowledge Graphs	Logical Consistency and World Knowledge	DeFacto, Kauwa-Kaate

social networks, especially on Twitter/X (Hui et al., 2018). It was developed by Indiana University as part of the OSoMe (Observatory on Social Media) project by the same researchers behind Botometer.

Multimodal. Multimodal approaches combine the analysis of visual and linguistic content, motivated by the fact that many fake news stories do not only consist of textual content, but are also image or video-based. Methods such as image forensics (e.g. reverse image search, manipulation detection) and multimodal models such as VisualBERT (Li et al., 2019) or VilBERT (Lu et al., 2019) are used to compare whether there is a contradiction between image content and text content. Both are multimodal transformer models designed to process text and image data together and being able to perform tasks such as a text-image coherence check or a multimodal sentiment analysis. The main difference between the two models lies in the architecture, particularly how image and text are combined.

Knowledge Graphs. Approaches for checking logical contradictions by comparison with structured knowledge databases (e.g. DBpedia, Wikidata) build knowledge graphs and check for logical consistency. Techniques include Triple Extraction (subject, predicate, object), Graph Reasoning or Knowledge Base Completion Consistency Check. DeFacto (Deep Fact Validation) was developed with the aim of automatically checking whether a claimed fact is true or false by comparing it with known facts from semantic knowledge databases (Linked Open Data, e.g. DBpedia) (Gerber et al., 2015). It provides additional evidence in text form, e.g. from Wikipedia paragraphs, for transparency and traceability. The research group continues to work on these topics and is expanding its model to include the validation of facts that are limited in time. The model thus takes into account points in time and validity periods and achieves greater accuracy (Qudus et al., 2023). The Kauwa-Kaate-System supports querying based on text as well as on images and video and can be accessed via WhatsApp or browser (Bagade et al., 2020). The input is compared with content collected from fact checking sites. In addition, a comparison is made with articles from established, trustworthy news sources.

2.3 Systems with Combined Approaches

The detection and evaluation of disinformation has long been studied in different ways with varying results. Many of the approaches presented above for evaluating an article have proven to be successful in their field. However, they only analyze a certain aspect of the article (e.g. sentiment, text style or source). It should be highlighted that some of the tools mentioned above are freely available and can be used and combined in external systems. It has been shown that the combination of different methods and approaches has led to more successful results. Also, the reliability of such tools depends to a large degree on the explainability of the rating results. Some innovative tools already combine a few techniques, e.g. XFake (Vosoughi et al., 2018), BRENDA (Botnevik et al., 2020) and Fake Tweet Buster (Saez-Trumper, 2014). Table 2 shows the combined approaches in the three systems in comparison to our platform.

The XFake system analyzes both attributes of the article (e.g. author, source) and statements. Three frameworks were developed for this purpose, with one designed for attribute analysis (corresponds to approach (2) Source Credibility), the second for the semantic analysis of statements (corresponds to approach (4) Sentiment Analysis) and the third for the linguistic analysis of statements (corresponds to approach (3) Style Analysis). In addition to the mere classification (fake vs. credible), XFake provides reasons about decisive factors (e.g. tense language, inconsistent sources), as well as relevant supporting examples and visualizations to facilitate interpretation. XFake was trained and evaluated on a dataset of thousands of political news claims verified by PolitiFact. The XFake tool considers already many analysis criteria (style, semantic and source analysis), but does not take into account the dissemination of an article. BRENDA recognizes and evaluates news sources (corresponds to approach (2) Source Credibility) and content (corresponds to approach (1) Fact Checking). They help users to distinguish trustworthy informa-

Table 2: Systems combining different approaches.

Approach	XFake	Brenda	FTP	Our tool
(1)		x		x
(2)	x	x	x	x
(3)	x			x
(4)	x			x
(5)			x	x
(6)			x	
(7)				

tion from disinformation through browser extensions or in social networks. It uses a tested deep neural network architecture in the background to automatically check facts and present the result with evidence to users by using an AI-classifier to assess whether a claim is likely to be true or false following a preliminary claim detection. At the same time, sources or evidence (e.g. fact check websites) are presented to confirm the claim. The Fake Tweet Buster (FTB) web application identifies tweets with fake images and users who regularly upload and/or spread fake information on Twitter/X by combining reverse image search (corresponds to approach (6) Multimodal), user analysis (corresponds to approach (2) Source Credibility) and crowd-sourcing (corresponds to approach (5) Social Signals) to detect malicious users. The FTB takes into account the dissemination of an article, but concentrates on a limited area of application: identifying tweets with fake content.

In our work, the focus is not on the mere reliability classification of an article. The tool developed by us is primarily intended to support the user in forming their own opinion on the trustworthiness of a news article. We follow an integrative approach by analyzing the article using different methods and presenting the results visually with suitable UI elements. In this way, we incorporate several of the aforementioned criteria (2, 3, 4, 5 and partly 1 claim detection) in a meaningful combination to provide the user with a comprehensive evaluation of the article. To the best of our knowledge, we are not yet aware of the existence of any other such instrument.

3 SYSTEM ARCHITECTURE

The system architecture of the developed platform with the technical system components involved and their interfaces to each other is outlined in Figure 1. A central controller coordinates all processes, addresses all required analysis components, and forwards data to the web interfaces. The controller also serves as an interpretation layer preparing the results of the individual services in such a way that they can be visually presented in an understandable form. The workflows from article analysis to the visual presentation of the results can be summarized as follows:

Web Scanner. The web service continuously searches for online sources on a topic defined by keywords and prepares the content found in a structured manner. The created article corpora can then be searched for a specific article URL. On request, all article features belonging to a URL are delivered to the central controller. In addition, a separate text extraction component is used to ensure an optimized and complete extraction of the entire body text of the article. For the prototype implementation, this text extraction component is limited to a list of predefined online sources (news publishers). Selected article features are prepared as part of the interpretation layer to be directly used in the visual presentation of the analysis results. This includes general article information (e.g. title, publication date, URL) as well as information on the dissemination of the article (referenced and referencing other articles).

Reliability-Rating. The component for analyzing article credibility assesses whether an article is more similar to online articles from credible sources or to articles from non-credible sources. For this evaluation, different feature domains are analyzed (text style, structure of the article website and article distribution network). The assessment of article credibility is based on an AI transformer model, which is pre-trained on the above-mentioned feature domains using extensive training data, including the FANG-COVID (Mattern et al., 2021) dataset. The credibility of the article sources in the training data is estimated using the NewsGuard rating (NewsGuard Technologies, 2025). The classification results of the feature domains are then comparatively weighted based on evaluated quality factors (f1-scores) and sent back to the controller as an overall assessment.

Claim Detection and Sentiment Analysis. A further analysis component breaks down the supplied article text into individual sentences. The individual sentences are further analyzed with regard to their sentiment (positive/negative) as well as whether the sentence represents a statement. The sentiment analysis is based on an AI transformer model pre-trained for the German language used to detect sentiment depending on the overall context of the sentence. The claim detection component uses a deep learning model that also uses AI transformer architectures. For the prototype implementation, the method was optimized for the detection of statements in the context of COVID-19 and MPox with the help of a manually annotated training data set. The individual results are also sent back to the controller at sentence level.

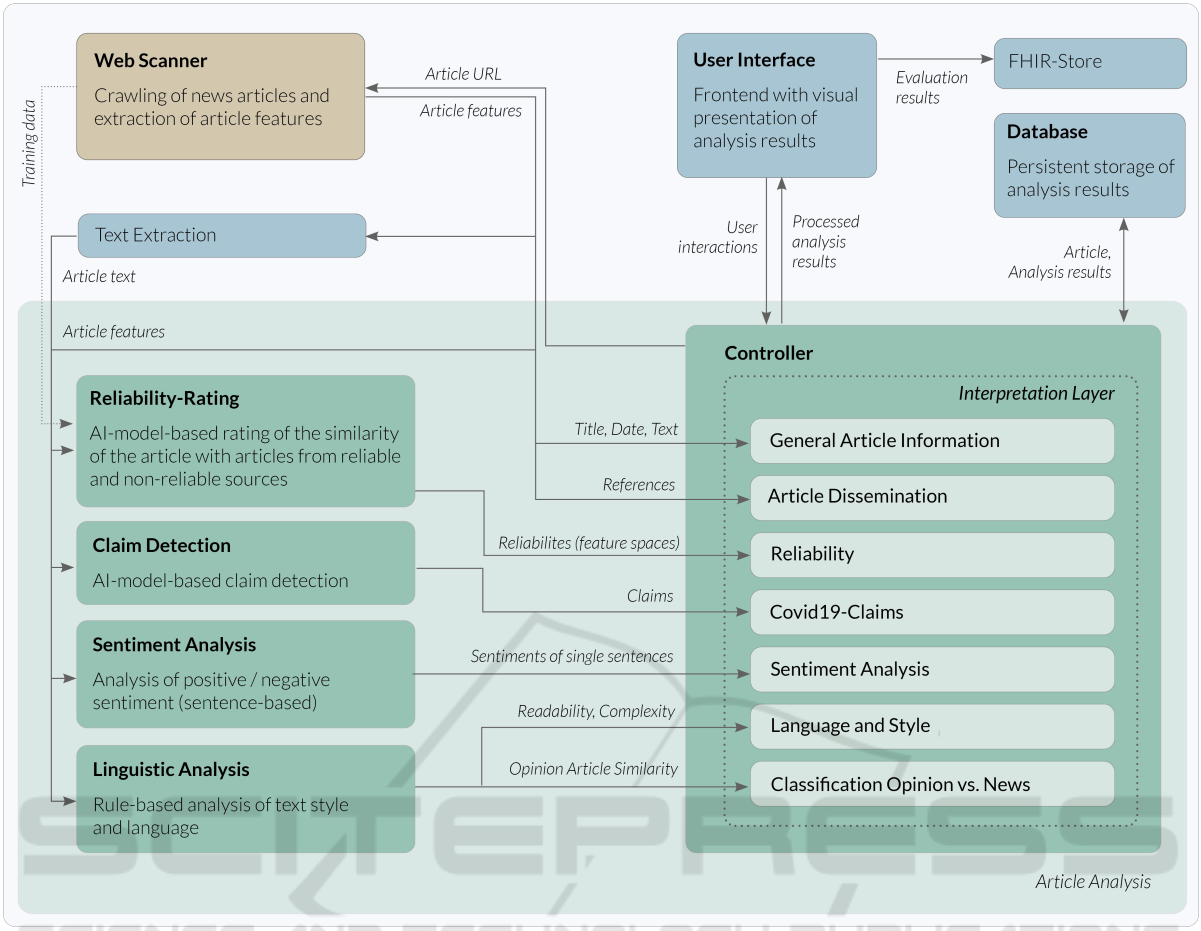

Figure 1: Technical system architecture of the implemented platform for assessing and visualizing online news reliability.

Linguistic Analysis. The extracted article text is also used for linguistic analyses. On the one hand, the text is evaluated in terms of lexical complexity (MTLD (McCarthy and Jarvis, 2010)) and readability (Flesch Reading Ease (Amstad, 1978)). On the other hand, a classification determines whether the text is more similar to an objective news article or to a subjective opinion article. The classification is based on the use of certain function words (pronouns, adverbs, etc.). With the help of a Support Vector Machine (SVM), which was trained with a corpus of 600 labeled news and opinion articles from 27 different publishers, similarity values of the text to opinion and news articles are calculated.

Database. The analysis results delivered to the controller are stored in a database for persistent storage. If an article URL to be analyzed already has analysis results stored, these are loaded directly from the database.

User Interface. The processed article features and analysis results are finally transferred to the user interface component (web frontend) visually presenting the results. Analogously, user actions on the web interface (e.g. entering a new article URL) are received by the controller and processed accordingly.

Finally, a FHIR store (Fast Healthcare Interoperability Resources) was designed as a backend component to store questionnaire-based result data as part of the evaluation (see chapter 5).

4 USER INTERFACE DESIGN

The graphical interfaces for presenting the analysis results were created through iterative UI design processes including multiple prototyping, testing and refinement steps. The final interfaces created are briefly presented and described below.

The mode of article input can be selected on the initial page. Either an existing URL in the database or the title and text for the article analysis can be entered. By clicking the button ‘article analysis’, a query is sent to the technical components (controller and database) and the article will be analysed on the

fly. Once the analysis has been completed, the article results are displayed in the user interface. In the section 'General Information' you find basic information (title, publication date, URL, source) on the analysed article, if available (Figure 2, upper part).

The 'Credibility' section displays the results whether the article is more similar to online articles from credible sources or more similar to articles from non-credible sources (Figure 2, lower part). For this assessment, the three feature areas 'Text style of the article', 'Distribution network' and 'Website structure of the article source' are analysed separately. The individual classification results of each category are then weighted comparatively and an overall summarised score on a scale of up to 100 is given. The external NewsGuard rating for the trustworthiness of the article source can also be viewed by clicking on 'Comparison with external source'.

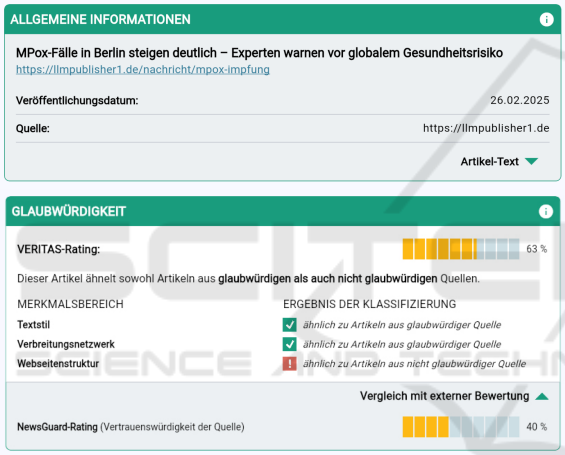

Figure 2: Visualization of general article information and results of reliability analysis.

In the 'Language and Style' section, the current article is analysed in terms of its readability, its complexity and its positive or negative sentiment. The results are presented on a scale from 1 (low) to 5 (high). A radar chart shows the correlation of these three text style areas (Figure 3, upper part). Here, the current article can also be compared with three different comparison corpora. The comparison corpora are subsets of the entire built-up article repository being formed from the training data set (FANG-COVID) or from articles previously analysed by the platform.

In the section 'Classification of Opinion vs. News Text' (Figure 3, lower part), a linguistic analysis using function words is carried out to classify to which extent this article (red dot) is more likely to be classified as a news article (green dots) or an opinion article (blue dots). The visualisation - reduced to two dimensions - illustrates the function word-based features of

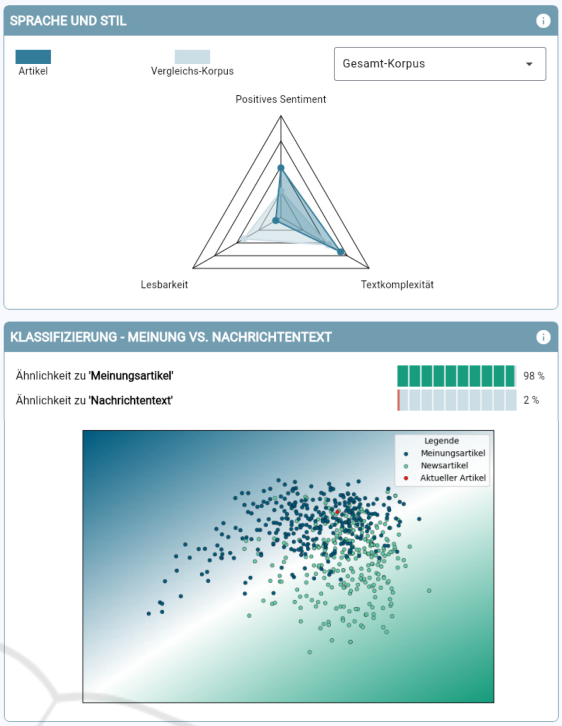

Figure 3: Visualization of 'Language and Style' analysis as well as results of the news and opinion article classification.

all texts on which the classification is based.

In the section 'Article Dissemination', the articles that refer to the analysed article (in-links) are listed in a bar chart chronologically (Figure 4). Also, the articles to which the analysed article refers (out-link) are listed. The visualisation also illustrates the media types (news portals, social media, other websites) of the in- and out-links indicating the media segments in which the article is distributed. If there are no further links in the article, the graphic remains empty.

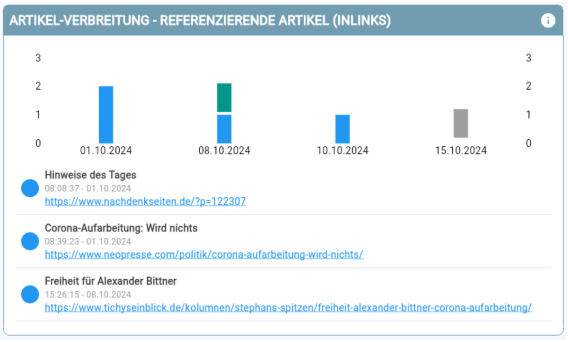

Figure 4: Presentation of article dissemination.

The section 'COVID-19 or MPox claims (Claim Detection)' identifies and highlights the claims or statements relating to COVID-19 or MPox in the analysed article at sentence level (Figure 5). An addi-

tional sentiment analysis colours individual sentences that convey a positive or negative sentiment (polarity) green or red. Beige stands for a neutral sentiment.

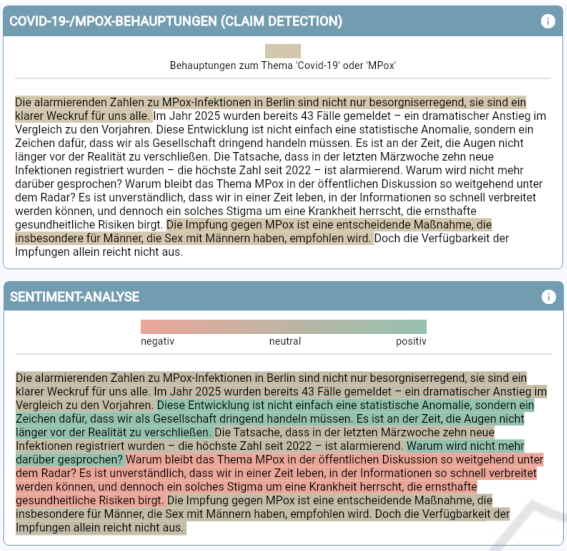

Figure 5: Visual output of components for claim Detection and sentiment analysis.

5 EVALUATION

An evaluation was conducted to find out how users perceive the tool by a quantitative and a qualitative section. The key findings are described below.

The application of the VERITAS demonstrator resulted in a predominantly positive overall evaluation. Even though the significance or usefulness of some displayed analysis criteria appeared less important according to the initial impressions of the participants, the tool generated a positive overall impression in terms of its presentation character – also with regard to the diversity of the displayed analysis criteria. The diverse analysis criteria presented are interesting, each with different potential: (1) The highly complex analysis of individual stylistic aspects at article level can be used to reveal potentially implicit structures (e.g., positive/negative sentiment) and to offer different approaches for interpreting the texts (opinionated vs. neutral). (2) The analysis of the article's context (e.g., dissemination) or of a fact seems to be an important criterion. This involves not only the direct linking of articles/posts to one another, but also the adoption and manipulation of a topic/fact from one article in other articles or posts (content similarity between articles/posts). (3) Certain stylistic feature spaces such as readability and complexity are considered less relevant for the assessment of the disinformative nature of an article.

A key finding was that it is difficult or impossible to clearly assess individual articles or excerpts from articles as credible ("not fake") or non-credible ("fake"). Automatically checking the truthfulness or reliability of content is very complex; even manually evaluating the truthfulness of individual statements is labor-intensive. Therefore, a tool like the VERITAS portal should rather be used to analyze and compare articles from different perspectives.

6 LIMITATIONS AND OUTLOOK

In terms of the variety of analysis criteria presented, the platform made a positive overall impression on users, although the significance and usefulness of individual selected parameters are not easy to understand at a first glance. However it turned out, that complex representations of analysis results, even if the single analysis parameters can be seen as easy to understand, lead to uncertainty, especially if a correlation of these parameters is visualised (e.g. with the radar chart). The work presented here provides a valuable starting point, but it would benefit from simplifying the interface and the generalization of claim detection through the addition of further training data. Certain areas of analysis, such as claim detection, are limited in their focus to the topics of COVID-19 and MPox.

The following findings can be concluded from the work on the project: It seems difficult to categorise an article as credible (non-fake) or not credible (fake) solely based on its textual characteristics. This is due to the fact that even less reputable portals and media often publish serious, factually neutral information within the article. According to the current state of the art, the truth or factuality of an article can only be checked in comparison with external sources that are as objective as possible. This requires the manual labelling of facts or objective statements by editors or fact checkers in order to ensure that the ground truth can be used for the further automated verification of the article's content. Automatically checking the truthfulness of article content is very complex, and even manually assessing the truthfulness of individual statements is time-consuming. A tool such as the presented platform should therefore not be used to assess the fake character (factuality) of individual articles, but merely to analyse articles from different perspectives and make them comparable. A good way of recognising to which extent an article has opinion-forming or manipulative intentions is by analysing the strength of opinion, sentiment and the credibility of

the source. Although the results of the analyses carried out for this purpose should not be presented in detail to not confuse users, the further technical development of a component for the evaluation of the strength of opinion of an article would be recommendable.

For more valid results, a large, unbiased new training data set with manually labeled manipulative and non-manipulative articles would therefore have to be set up in further developments. Analysing article sources (e.g. news publishers) and the article distribution network (especially the distribution in various social networks) is also very important and should be expanded in follow-up projects. In particular, the dissemination analysis should extract facts within the article text and link them to official announcements and legislative texts similar to the Knowledge-Graph based approach (e.g. press releases, legal texts, if necessary also complete original quotations) in order to show the provenance of the information and any alienating editing strategies.

In the future, we see further potential applications on the presented platform like searching reliable sources on specialised topics and checking the citation of references for scientific papers, as well as a 'neutrality analysis' of one's own texts.

REFERENCES

Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen?* Studenten-Schreib-Service.

Bagade, A., Pale, A., Sheth, S., Agarwal, M., Chakrabarti, S., Chebrolu, K., and Sudarshan, S. (2020). The kauwa-kaate fake news detection system. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 302–306.

Botnevik, B., Sakariassen, E., and Setty, V. (2020). Brenda: Browser extension for fake news detection. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 2117–2120.

BR24 (2025). Faktenfuchs Faktenchecks. https://www.br.de/nachrichten/faktenfuchs-faktencheck.

CORRECTIV (2025). Faktenforum. https://www.faktenforum.org/.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.

FullFact (2025). Full Fact. https://fullfact.org/.

Gerber, D., Esteves, D., Lehmann, J., Bühmann, L., Usbeck, R., Ngomo, A.-C. N., and Speck, R. (2015). Defacto—temporal and multilingual deep fact validation. *Journal of Web Semantics*, 35:85–101.

Hartwig, K. and Reuter, C. (2019). Trustytweet: An indicator-based browser-plugin to assist users in dealing with fake news on twitter.

Hui, P.-M., Shao, C., Flammini, A., Menczer, F., and Ciampaglia, G. L. (2018). The hoaxy misinformation and fact-checking diffusion network. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.

IFCN (2025). International Fact-Checking Network - Poynter. https://www.poynter.org/ifcn/.

Li, L. H., Yatskar, M., Yin, D., Hsieh, C.-J., and Chang, K.-W. (2019). Visualbert: A simple and performant baseline for vision and language. *arXiv preprint arXiv:1908.03557*.

Lu, J., Batra, D., Parikh, D., and Lee, S. (2019). Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems*, 32.

Mattern, J., Qiao, Y., Kerz, E., Wiechmann, D., and Strohmaier, M. (2021). Fang-covid: A new large-scale benchmark dataset for fake news detection in german. In *Proceedings of the fourth workshop on fact extraction and verification (fever)*, pages 78–91.

McCarthy, P. M. and Jarvis, S. (2010). Mtld, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2):381–392.

NewsGuard Technologies (2025). News reliability ratings. https://www.newsguardtech.com/solutions/news-reliability-ratings/.

Poynter (2025). Politifact. https://www.politifact.com/.

Qudus, U., Röder, M., Kirrane, S., and Ngomo, A.-C. N. (2023). Temporalfc: a temporal fact checking approach over knowledge graphs. In *International Semantic Web Conference*, pages 465–483. Springer.

Saez-Trumper, D. (2014). Fake tweet buster: a webtool to identify users promoting fake news ontwitter. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 316–317.

Shu, K., Cui, L., Wang, S., Lee, D., and Liu, H. (2019). defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405.

Tian, L., Zhang, X., and Peng, M. (2020). Fakefinder: twitter fake news detection on mobile. In *Companion Proceedings of the Web Conference 2020*, pages 79–80.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *science*, 359(6380):1146–1151.

Yang, K.-C., Ferrara, E., and Menczer, F. (2022). Botometer 101: Social bot practicum for computational social scientists. *Journal of computational social science*, 5(2):1511–1528.

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., and Choi, Y. (2019). Defending against neural fake news. *Advances in neural information processing systems*, 32.