

Leveraging Large Language Models for News Values Analysis (Extended Abstract)

Gullal S. Cheema^{1,*}, Massiollah Azimi², Ralph Ewerth^{1,3} and Eric Müller-Budack³

¹L3S Research Center, Leibniz University Hannover, Germany

²Leibniz University Hannover, Germany

³TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

Abstract

This study explores the use of large language models (LLMs) for the automated extraction of news values, marking the first effort to apply LLMs to this task. We evaluate open-source models, including LLaMA3, Yi-1.5, and Qwen2, using structured prompts to detect four news values: sentiment, geographical proximity, timeliness, and eliteness. Results demonstrate promising performance with few-shot prompting over zero-shot evaluation. Manual annotations revealed challenges due to the complexity and subjectivity of news values, with moderate inter-annotator agreement (Cohen's kappa of 0.41), emphasizing the need for larger datasets and multiple annotators to ensure reliable ground truth labels. Fake news analysis showed distinct patterns, including a higher emphasis on negativity and prominence compared to true news.

Keywords

News values detection, news analysis, generative AI, large language models

1. Motivation

News values, as introduced by Galtung and Ruge [1], capture the attributes that determine the "news-worthiness" of actors, events, and issues. They guide how news media construct narratives and news-worthiness [1], shaping public perception and discourse. Contemporary studies have expanded their applicability to diverse platforms, from social media engagement [2] to the analysis of fake news [3], underscoring their relevance in traditional and digital contexts. For instance, Tandoc et al. [3] highlighted distinct patterns in fake news, identifying a prevalence of *timeliness*, *negativity*, and *prominence*, alongside a lack of objectivity. These approaches highlight the growing need for scalable, context-aware computational methods, such as large language models (LLMs), to analyze and extract these values from news articles effectively.

So far, very few computational models on the detection of news values in news articles have been introduced. Existing approaches have employed methods such as statistical features, linguistic analysis, and machine learning approaches like Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs) to extract and classify news values. Studies like Potts et al. [4] and Bednarek et al. [5] utilized these methods on specific corpora, while others, such as di Buono et al. [6] and Piotrkowicz et al. [7], incorporated word embeddings and sentiment analysis. While these efforts provided interesting insights, they rely on rather outdated AI approaches and have not yet leveraged powerful, generative AI approaches [8] to detect news values.

Disinformation, Misinformation and Learning in the Age of Generative AI: Joint Proceedings of the 1st International Workshop on Disinformation and Misinformation in the Age of Generative AI (DISMISS-FAKE'25) and the 4th International Workshop on Investigating Learning during Web Search (IWILDS'25) co-located with 18th International ACM WSDM Conference on Web Search and Data Mining (WSDM 2025)

*Corresponding author.

✉ gullal.cheema@tib.eu (G. S. Cheema); massiollah.azimi@stud.uni-hannover.de (M. Azimi); ralph.ewerth@tib.eu (R. Ewerth); eric.mueller@tib.eu (E. Müller-Budack)

ORCID 0000-0003-4354-9629 (G. S. Cheema); 0000-0003-0918-6297 (R. Ewerth); 0000-0002-6802-1241 (E. Müller-Budack)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Classification of News Values using LLMs

This work addresses the aforementioned limitations, and makes three primary contributions towards the use of LLMs for computational news values analysis.

(1) *Evaluation of Open-Source LLMs*: We evaluate the performance of state-of-the-art open-source LLMs, including *LLaMA3* [9], *Yi-1.5* [10], and *Qwen2* [8], in detecting news values. These include *sentiment* (positive or negative tone), *geographical proximity* (local, regional, national, or global relevance), *timeliness* (old, recent, ongoing, or future events), and *eliteness* (presence of influential individuals or organizations). Definitions and subcategories of news values are adapted from Cheema et al. [11]. To optimize model performance, we use structured prompt design with clear context, task definition, and output constraints. We compare structured JSON outputs (Table 2) with unstructured formats to assess consistency and evaluate the impact of zero-shot and few-shot prompting (in-context learning) on annotation accuracy. Structured outputs consistently outperform unstructured ones; therefore, we present comparison results and prompt types exclusively for structured outputs in Table 1. The evaluation is conducted separately for each news value.

(2) *LLM vs Human Annotation*: We compare LLM-generated annotations with human annotations to evaluate strengths and limitations. Two annotators, each with a computer science background, label 20 news articles (10 true, 10 fake) from the *ISOT Fake News Dataset* [12], which includes true news articles from *Reuters* and fake news flagged by *Politifact*. Annotation guidelines are adapted and refined from Cheema et al. [11], ensuring a systematic and consistent evaluation of annotation quality.

(3) *True vs. Fake News Analysis*: Leveraging the best performing LLM from the previous step, we analyze 100 news articles (50 true, 50 fake) to examine differences in detected news value patterns. Through frequency and bigram analysis, we identify distinctive characteristics in true and fake news, providing insights into how these patterns may contribute to automated misinformation detection.

These contributions lay the groundwork for future research in automating news values analysis and applying it to diverse journalistic and computational contexts.

3. Preliminary Results

Inter-Annotator Agreement: To validate the reliability of manual annotations, we compared the annotations made by two human annotators using Cohen’s kappa. The average kappa score across 20 news articles was 0.41, indicating a moderate agreement. Individual kappa values varied widely, ranging from 0.13 to 0.76, reflecting the complexity and subjectivity inherent in annotating news values despite consistent guidelines.

LLM performance: *Qwen2* [8] achieved the best results across the evaluated LLMs. Results for all LLMs are provided in Table 1. Few-shot prompting with two examples, consistently outperformed zero-shot, with notable improvements across all models, particularly for complex news values like *timeliness* and *geographical proximity*. *Qwen2* delivered robust results across most news values but struggled with *sentiment* in the zero-shot setting. Prompt complexity (Table 2) also influenced outcomes, with simple

Table 1

F1-scores of three models (*Qwen2* [8], *Llama-3* [9], and *Yi-1.5* [10]) for detection of news values using ground truth data from two annotators (A1, A2). **Bold** denotes best, and dashed underline denotes second-best performance.

News Value	Qwen2				Llama-3				Yi-1.5			
	Zero-shot		Few-shot		Zero-shot		Few-shot		Zero-shot		Few-shot	
	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2
Geographical Proximity	0.25	0.27	0.63	0.72	0.00	0.00	0.63	0.60	<u>0.08</u>	0.18	<u>0.47</u>	0.63
Timeliness	0.40	0.35	0.67	0.43	0.00	0.00	<u>0.59</u>	<u>0.37</u>	<u>0.16</u>	0.00	0.31	0.16
Eliteness	<u>0.21</u>	<u>0.29</u>	0.64	0.72	0.05	0.15	0.31	0.39	0.31	0.42	<u>0.35</u>	0.46
Sentiment	<u>0.04</u>	<u>0.03</u>	0.53	0.55	0.01	0.01	0.33	0.39	0.26	0.29	<u>0.51</u>	0.55

output format (Type 1) for direct classification excelling in zero-shot and sentence-wise analysis (Type 3) with justification format yielding better results in the few-shot setup.

Table 2

Overview of prompt types. “FS” indicates few-shot usage, which includes one or more examples. Each instruction includes a brief description of the news value with optional additions per prompt type.

Feature	Type 1	Type 2	Type 3
Instruction	Minimal, direct	Asks for justification	Sentence-level analysis + justification
Definition	Basic within instruction	Basic within instruction	Extended detailed definition
Categories	Included	Included	Included
Output Format	Basic JSON (text_span, label, start, end)	Basic JSON + includes justification	Basic JSON per-sentence + justification
Examples	Provided when FS	Provided when FS	Provided when FS
Key Focus	Direct classification	Reasoning behind classification	Fine-grained, sentence-level analysis

Comparison of News Values between True and Fake News: The analysis revealed distinct differences in narrative patterns between true and fake news articles. Mentions of prominent individuals and organizations were frequent in both categories, reflecting the political focus of the dataset. Events tied closely to the timing of publication were common, while references to older or future events appeared less frequently, aligning with the topical nature of political reporting. A notable finding was the significantly higher emphasis on negative language in fake news—over three times more frequent than in true news—suggesting a strategic emphasis on negativity to attract attention or shape perceptions. Additionally, patterns in news value bigrams revealed that fake news often amplifies *eliteness* and *negative sentiment* through repeated associations, such as consecutive mentions of *influential entities* or chains of *negative sentiment*. These results suggest that fake news leverages emotional and influential elements [3] to serve political or ideological purposes, offering valuable insights for detecting misinformation.

4. Conclusion and Future Work

This work presents the first study leveraging LLMs for automated extraction of news values, demonstrating promising results on a small dataset. Manual annotations revealed challenges with complexity and subjectivity, as shown by moderate inter-annotator agreement (Cohen’s kappa of 0.41), highlighting the need for larger datasets and more annotators to establish reliable ground truth labels. Future work will expand to more news values, larger datasets, and diverse media contexts. Combining manual and LLM-generated annotations for semi-supervised datasets and exploring efficient fine-tuning methods, such as instruction-tuning, could enhance performance.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF, FakeNarratives project, no. 16KIS1517).

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4o in order to: Paraphrase and reword, Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] M. Bednarek, H. Caple, *The Discourse of News Values: How News Organizations Create Newsworthiness*, Oxford University Press, 2017. URL: <https://doi.org/10.1093/acprof:oso/9780190653934.001.0001>. doi:10.1093/acprof:oso/9780190653934.001.0001.
- [2] T. Araujo, T. G. van der Meer, News values on social media: Exploring what drives peaks in user activity about organizations on twitter, *Journalism* 21 (2020) 633–651.
- [3] E. Tandoc, R. Thomas, L. Bishop, What is (fake) news? analyzing news values (and more) in fake stories. *media and communication*, 9 (1), 110–119, 2021.
- [4] A. Potts, M. Bednarek, H. Caple, How can computer-based methods help researchers to investigate news values in large datasets? a corpus linguistic study of the construction of newsworthiness in the reporting on hurricane katrina, *Discourse & Communication* 9 (2015) 149–172.
- [5] M. Bednarek, H. Caple, C. Huan, Computer-based analysis of news values: A case study on national day reporting, *Journalism Studies* 22 (2021) 702–722.
- [6] M. P. di Buono, J. Snajder, B. D. Basic, G. Glavas, M. Tutek, N. Milic-Frayling, Predicting news values from headline text and emotions, in: *Proceedings of the 2017 Workshop: Natural Language Processing meets Journalism, NLPmJ@EMNLP*, Copenhagen, Denmark, September 7, 2017, Association for Computational Linguistics, 2017, pp. 1–6. doi:10.18653/V1/W17-4201.
- [7] A. Piotrkowicz, V. Dimitrova, K. Markert, Automatic extraction of news values from headline text, in: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017*, Valencia, Spain, April 3–7, 2017, Student Research Workshop, Association for Computational Linguistics, 2017, pp. 64–74. doi:10.18653/V1/E17-4007.
- [8] A. Yang *et al.*, Qwen2 technical report, *CoRR abs/2407.10671* (2024). doi:10.48550/ARXIV.2407.10671. arXiv:2407.10671.
- [9] A. Dubey *et al.*, The llama 3 herd of models, *CoRR abs/2407.21783* (2024). doi:10.48550/ARXIV.2407.21783. arXiv:2407.21783.
- [10] A. Young *et al.*, Yi: Open foundation models by 01.ai, *CoRR abs/2403.04652* (2024). doi:10.48550/ARXIV.2403.04652. arXiv:2403.04652.
- [11] G. S. Cheema, S. Hakimov, E. Müller-Budack, C. Otto, J. A. Bateman, R. Ewerth, Understanding image-text relations and news values for multimodal news analysis, *Frontiers Artif. Intell.* 6 (2023). doi:10.3389/FRAI.2023.1125533.
- [12] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classification, *Security and Privacy* 1 (2018) e9.