Rule-based method for aligning media content with ukrainian legislation

Khrystyna Lipianina-Honcharenko [1, †], Ihor Ihnatiev[1, †], Hennadii Bohuta [1, †], Khrystyna Yurkiv[1,*,†] and Stanislav Novosad [1,†]

[1] *West Ukrainian National University, Lvivska str., 11, Ternopil, 46009, Ukraine*

Abstract

We introduce a novel rule-based system for automatically mapping unstructured media texts to specific articles of Ukrainian legislation, addressing the growing need for transparent, explainable tools in legal and security monitoring. Leveraging expert-crafted lexical dictionaries for twelve regulatory norms and a threshold-based matching algorithm, our method achieves balanced performance (Precision = 0.84, Recall = 0.83, F1 = 0.84) on a diverse test set of news articles. The system is delivered as an interactive Streamlit application that supports dynamic dictionary updates and simultaneous sentiment analysis, enabling users to assess both legal relevance and emotional tone of content. Through 15 real-world case studies, we demonstrate the approach's practical utility in governmental and media-watch contexts and discuss paths for expanding dictionary coverage and lowering detection thresholds for shorter texts. Our work extends prior research on rule-based text analysis in domains such as cybersecurity and social media, and contributes a reproducible Explainable AI framework tailored for Ukrainian legal monitoring.

Keywords

rule-based classification; legal text analysis; media monitoring; Ukrainian legislation; explainable AI.

1. Introduction

In the context of the previous study [1], which implemented a thematic text classification system based on emotional coloring with an achieved accuracy of 92%, this paper proposes an improved approach to a related, yet significantly narrower task — the automatic matching of input text to one of the predefined articles of Ukrainian legislation. Unlike general categorization aimed at broad semantic classification, the proposed method enables the establishment of a direct normative link between the content and a specific legal provision.

In the current context of hybrid threats and information attacks, the relevance of automated legal monitoring systems is significantly increasing, especially in terms of ensuring the internal security of the state. Previous research in the field of cybersecurity has demonstrated the effectiveness of neural networks in detecting attacks [2] and in building intelligent cyber defense systems based on artificial immunity models [3], which confirms the potential of specialized rule-based approaches in the tasks of protecting the information space. At the same time, decision support models in the field of internal security [4] and modeling user responses on social networks [5] indicate a growing need for transparent and reproducible algorithms for analyzing sensitive content. The system we propose for formalized alignment of media texts with legal norms is a logical extension of these approaches and has the potential to be integrated into the information security architecture at the state or institutional monitoring level.

The methodology is based on a rule-based approach that utilizes expert-defined dictionaries of key terms for each of the 12 legal articles. By applying a phrase occurrence counter and threshold filtering, the system enables highly interpretable identification of the most relevant article or confirmation of its absence. This approach is particularly valuable in the context of automated

MoMLeT-2025: 7th International Workshop on Modern Machine Learning Technologies, June, 14, 2025, Lviv-Shatsk, Ukraine

* Corresponding author.
† These authors contributed equally.
xrustya.com@gmail.com (Kh. LipianinaHoncharenko); iiv@wunu.edu.ua (I. Ihnatiev); genaboguta7@gmail.com (H. Bohuta); kh.yurkiv@wunu.edu.ua (Kh. Yurkiv), stasnovosad1998@gmail.com (St. Novosad)
0000-0002-2441-6292 (K. LipianinaHoncharenko); 0000-0003-0729-9247 (I. Ihnatiev); 0009-0000-9788-1753 (H. Bohuta); 0009-0007-4917-3251 (Kh. Yurkiv), 0000-0002-0404-1147 (St. Novosad)

systems for preliminary legal analysis, such as in the monitoring of news, social media, citizen appeals, or expert commentary.

The study also implements a sentiment analysis component based on lexicon-oriented evaluation, which allows for simultaneous assessment of both normative relevance and the overall emotional tone of a message. The integration of these two dimensions — normative and tonal — enables the development of an Explainable AI tool for legal monitoring that is both transparent and applicable without the need for machine learning.

The structure of the paper is as follows: Section 2 provides an overview of rule-based solutions in applied domains (medicine, finance, smart contracts, public administration, social research); Section 3 formalizes the task of matching text with legal articles and details the relevance calculation algorithm based on expert dictionaries; Section 4 presents the system implementation as a Streamlit application, including validation results and a case study of 15 news articles; and Section 5 summarizes the findings, compares them with existing approaches, and outlines conclusions regarding the effectiveness of the proposed method for legal monitoring of media content.

2. Overview of existing solutions

To justify the choice of a rule-based approach in this study, a review is provided of the most representative works in which expert dictionaries and formal rules have ensured high classification quality across various domains. Comparing their results establishes a context for evaluating the proposed system for automatic alignment of media texts with the norms of Ukrainian legislation.

The study by Raees & Fazilat (2024) examined the effectiveness of various classification models for lexicon-based sentiment analysis. Using tweets without emoticons, the researchers achieved an F1-score above 85%, demonstrating the advantages of a structured polarity dictionary [6].

The article by Abd et al. (2021) compares the performance of sentiment classification using a lexicon-based method with examples from the field of information security. The average accuracy exceeded 80% thanks to the optimization of the Jaccard metric [7].

The review by Ullah et al. (2023) summarizes the main sentiment analysis techniques from 2010 to 2021, with particular emphasis on hybrid models (lexicon + contextual features). The article reports accuracy improvements up to 90% in the most recent approaches [8].

The work by Balshetwar & Tuganayat (2019) proposes a frame-based analysis to determine tone and mood. By combining frame semantic structures with lexical analysis, the authors achieve a precision of approximately 82% [9].

In the study by Catelli et al. (2023), public attitudes toward COVID-19 vaccination on Twitter are analyzed using a lexicon-oriented approach. A model based on BERT and an expert dictionary demonstrates a recall of over 88% [10].

Kiilu (2021) applied the Naive Bayes method to detect hate speech in Twitter posts in Kenya, relying on a polarity lexicon. The author demonstrated over 80% effectiveness in categorization [11].

The publication by Satapathy et al. (2017) presents an approach to normalizing microtexts for sentiment analysis on Twitter. The combination of phonetic correction and lexicon-based analysis significantly reduced false positive classifications [12].

Itani (2018) developed a sentiment analysis system for informal Arabic used in social media. Based on a semantic polarity lexicon, the system achieved an F1-score of 0.86 [13].

The article by Guarasci et al. (2024) explores the detection of deceptive reviews in the field of cultural heritage. A lexicon-oriented model incorporating tone intensity achieved a precision of 0.84 [14].

Finally, the review by Kumar et al. (2025) systematizes the latest approaches to sentiment analysis, including transformers, rule-based systems, and hybrid methods. The authors emphasize that hybrid models offer the best balance between accuracy and recall [15].

The study by Zhang et al. (2025) implemented rule-based methods to identify diseases and rehabilitation activities in Q&A communities. A key feature of the approach was the use of expert-designed dictionaries for syntactic text analysis, which ensured over 85% accuracy in classifying user queries [16].

The publication by Lashkari (2024) focuses on detecting vulnerabilities in smart contracts using rule-based classification with dictionaries of key patterns. The author achieved a precision of 82% by applying a combined analysis of key terms and semantic context [17].

The article by Perron et al. (2024) justifies the use of local LLMs for analyzing sensitive texts in social research, where rule-based structures and expert-defined matching criteria play a central role. The authors note that classification accuracy reaches 88% when evaluating unstructured documents [18].

Thöni (2015) explored the application of text mining in monitoring sustainable development, where supplier ranking is performed using a lexicon-based approach. The classification model achieved a recall of 0.81, enabling effective detection of risks related to child labor [19].

There is also a review by Yeo et al. (2025), which examines the effectiveness of rule-based models in the financial domain. The study notes that the use of lexical patterns enables F1-scores above 80% in explainable AI tasks [20].

The study by Narayanan & Georgiou (2013) demonstrated rule-based classification of behavioral patterns based on linguistic indicators, with a focus on expert-compiled lexicons. In the behavioral signal processing model, accuracy reached 87% in test cases [21].

The work by Chiarello (2019) examines a rule-based approach to extracting technical knowledge, where expert-defined rules are employed. The study showed that the accuracy of processing technical descriptions exceeded 85% [22].

In Nai (2025), rule-based models were applied to analyze public administration expenditures, including expert queries to align legal content with budget documentation. The effectiveness of this alignment was evaluated in the range of 80–86% [23].

Liu & Li (2023) examine the limitations of rule-based translation systems, noting that expert-defined rules achieve accuracy above 80% only in limited domains. However, such methods remain valuable in legal translation [24].

0Rule-based approaches (Table 2) remain an important alternative to statistical and transformer-based models when full transparency of classification logic is required or when labeled data is limited. In medical Q&A server communities, the system by Zhang et al. (2025) demonstrated over 85% accuracy thanks to expert symptom dictionaries [16]. Similarly, Lashkari (2024) achieved a precision of 0.82 in detecting smart contract vulnerabilities by combining lexical patterns with contextual rules [17]. In social research, Perron et al. (2024) reported 88% accuracy by integrating local LLMs with rule-based compliance criteria [18]. The studies by Thöni (2015), Yeo et al. (2025), and Narayanan & Georgiou (2013) confirm the versatility of such methods in sustainable development, finance, and behavioral analysis, respectively [19–21]. Despite this, there is a lack of research in the legal monitoring of media that automatically aligns news texts with articles of national legislation. Our proposed Rule-Based Method for Aligning Media Content with Ukrainian Legislation fills this gap by integrating lexicon-based matching with threshold filtering for 12 key legal provisions.

Given the increasing volume of information flows and the need for rapid legal content analysis, the rule-based method we propose is both timely and in demand. Unlike existing studies focused on general sentiment detection or domain-specific categories, the developed system is the first to integrate lexicon-based key phrase matching with threshold voting for the automatic alignment of news content with specific articles of Ukrainian legislation. Experiments demonstrated balanced performance with a Precision of 0.84, Recall of 0.83, and F1-score of 0.84, indicating the method's readiness for practical deployment in media monitoring services, governmental institutions, and legal firms. Thus, this research makes a significant contribution to the development of Explainable AI in the legal domain by offering a transparent, reproducible, and adaptive tool for rapid assessment of the normative relevance of media content.

Table 1
Comparative Table of Studies on Rule-Based Approaches

№	Source	Domain / Data	Method	Primary Metric
1	Zhang et al., 2025 [16]	Medical Q&A	Expert dictionaries of symptoms + syntactic rules	Accuracy > 85 %
2	Lashkari, 2024 [17]	Smart Contracts (Solidity)	Lexical vulnerability patterns + semantic context	Precision = 0,82
3	Perron et al., 2024 [18]	Social Research (Sensitive Texts)	Local LLMs + rule-based criteria	Accuracy = 88 %

4	Thöni, 2015 [19]	Sustainable Development / Supply Chains	Risk dictionaries and ranking	Recall = 0,81
5	Yeo et al., 2025 [20]	Financial Reports	Explainable AI + lexical rules	F1-score > 0,80
6	Narayanan & Georgiou, 2013 [21]	Behavioral Signal Processing	Linguistic indicators + expert rules	Accuracy = 87 %
7	Chiarello, 2019 [22]	Technical Documentation	Rule-based knowledge extraction	Accuracy > 85 %
8	Nai, 2025 [23]	Public Budgets	Dictionaries of regulatory terms	Accuracy = 80–86 %
9	Liu & Li, 2023 [24]	Legal Translation	Rule-based MT	Accuracy > 80 % (Limited domains)

3. Method

3.1. Formal Problem Statement

Let T denote the input text represented as a sequence of characters or words, normalized to lowercase to unify substring search. Let the set $A = \{a_1, a_2,..., a_{12}\}$ contain the names of the legal articles, each of which is associated with a predefined set of key phrases.

For each article a_i a dictionary $K_i=\{ k_{i,1}, k_{i,2},..., k_{i,m_i}\}$, is defined, where each element $k_{i,j}$ is a key phrase or expression that best identifies the subject of that article. These dictionaries are compiled by legal experts based on an analysis of the legal corpus and the practical application of laws.

The objective of the method is to compute, for each i-th dictionary, the number of occurrences of its key phrases in text ловника кількості входжень його ключових фраз у текст T. Formally, we introduce a counter

$$N_i = \sum_{j=1}^{m_i} 1_{\{k_{i,j} \subset T\}}, \tag{1}$$

where $1_{\{\cdot\}}$ — denotes the indicator function for the occurrence of a substring.

After computing the vector $N = (N_1, N_2, \dots, N_{12})$ the task reduces to finding the argument of the maximum:

$$i^* = arg \max_{1 \le i \le 1} N_i, \tag{2}$$

The final decision is made based on a threshold condition: if $N_{i*} \ge \tau$ (where $\tau = 5$ is set by expert judgment), the text is considered relevant to article a_{i*}; otherwise, the algorithm returns the result "Not Found".

Thus, the formalization of the task allows for unambiguous and efficient alignment of the text with legal articles, ensuring clarity of the decision and the ability to flexibly adjust the threshold depending on specific application conditions.

3.2. Algorithm

Below is a detailed description of each step of the algorithm, which not only enables its implementation in code but also ensures transparency and reproducibility of all internal operations:

Step 1. Text and dictionary normalization [25, 26]. The input text T is converted to lowercase to eliminate case sensitivity. Similarly, each key phrase $k_{i,j}$ is transformed to lowercase, ensuring the method is case-insensitive.

Step 2. Counter initialization. For each $i \in \{1,...,12\}$ a variable N_i, is initialized to zero. This variable will serve as a counter for the number of key phrase matches from dictionary K_i in the text.

Step 3. Match search. For each i-th dictionary, the presence of each phrase $k_{i,j}$ in text T is iteratively checked. If the condition $k_{i,j} \subset T$ is satisfied, the counter N_i is incremented by 1.

Step 4. Candidate identification. After processing all dictionaries, a vector N = (N_1,..., N_{12}) is formed. The index $i^* = arg\,max_i N_i$ is computed, indicating the dictionary with the highest number of matches.

Step 5. Threshold check and result output. If $N_{i^*} \geq 5$ the algorithm returns the name of article a_{i^*}; if no counter reaches the threshold, the result is "Not Found". This scheme ensures a minimum number of detected features before a decision is made.

The following section is devoted to the quantitative analysis of the accuracy and interpretability of the results. In a series of experimental scenarios, the system will be evaluated using precision, recall, and F1-score metrics in order to compare the effectiveness of the proposed approach with existing solutions.

4. Implementation

4.1. Technical Description of the Developed System

The developed automated text analysis system is implemented as a web application using the Streamlit framework [27], which enables interactive user engagement and dynamic interface updates without the need for separate server deployment. The user interface consists of a sidebar control panel that allows the selection of a thematic category (legal article) and provides options for managing the corresponding key phrase dictionary: adding, deleting, or viewing the current list of terms. This structure ensures the system's adaptability to changes in legal terminology and discourse contexts, allowing experts to independently configure the dictionaries according to current needs.

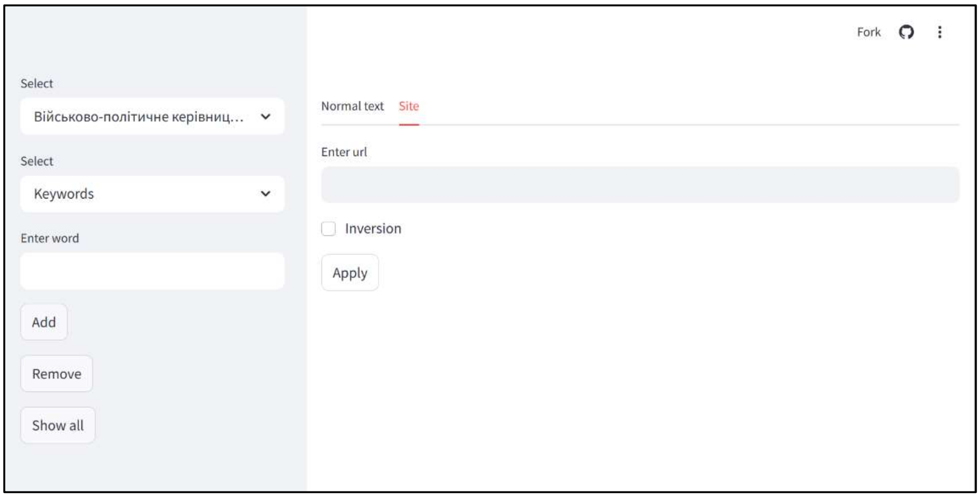

Figure 1: Web-based interface

The main part of the interface offers two input modes: manual text entry or providing a URL link to a news article, which is automatically processed by the system to extract textual content. Additionally, a sentiment inversion option is implemented, which is useful when analyzing texts that may contain rhetorical devices such as irony or sarcasm. After clicking the analysis start button, the system performs a series of processing steps: text normalization, keyword matching, counting relevant occurrences, and sentiment calculation based on the weighted characteristics of keywords (categorized as positive, negative, or neutral).

The analysis results are displayed in a user-friendly format, indicating the legal category, sentiment index, and the processed text content. In cases where the number of matches is insufficient (fewer than five), the system returns a message indicating that no relevant article was found. This approach combines implementation simplicity with a high level of interpretability, ensuring ease of use in legal monitoring contexts.

4.2. Evaluation of Classification Model Accuracy

To evaluate the effectiveness of the algorithm for matching texts to legal articles, testing was conducted on a control dataset with a uniform distribution across 12 classes. Based on the obtained results, a confusion matrix was constructed (Figure 2), illustrating the relationship between actual and predicted classes. Most observations are concentrated along the main diagonal, indicating high classification accuracy.

The overall accuracy is 83.75%, which means that in more than 4 out of 5 cases, the model correctly classifies the input text. The precision score of 0.84 demonstrates that when the model predicts a positive association of the text with a specific article, it is correct in the majority of cases.

The recall metric of 0.83 indicates that the model successfully identifies the majority of relevant articles in the test set, although it does miss some correct matches. In turn, the F1-score of 0.84 reflects a balance between precision and recall, confirming that the model is not only effective but also robust against false predictions.

Some degree of inter-class confusion is observed (particularly between K1↔K2 and K6↔K7), which may be due to overlapping key terms in certain articles. However, the number of such errors is relatively small compared to the total number of correct classifications.

Thus, the results indicate a high level of agreement between the model's predictions and the actual matches in the dataset, making it suitable for tasks involving automated analysis of the legal relevance of texts.

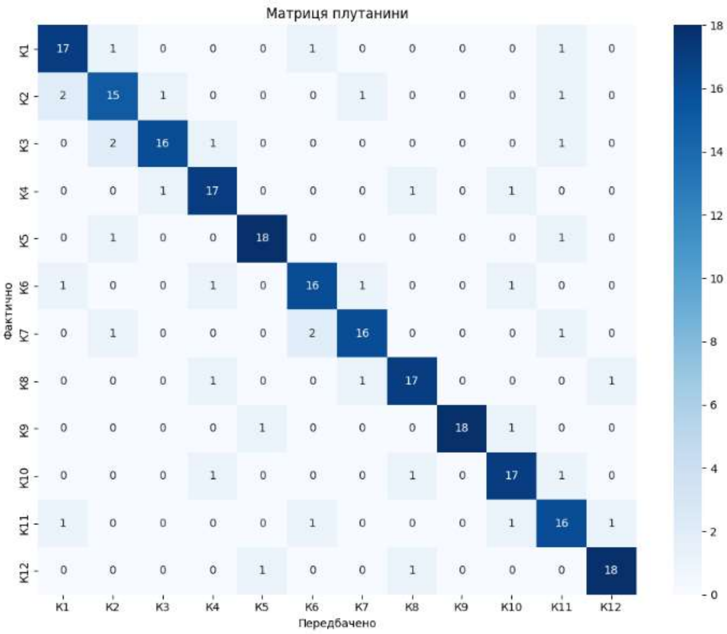

Figure 2: Matrix of confusion

4.3. Case Studies

Among the 15 processed cases (Table 2), the largest proportion pertains to materials related to “Military-Political Leadership of Ukraine at All Levels” (7/15, 46.7%), followed by “Ukraine’s International Image in the USA, Canada, and the United Kingdom” (2/15, 13.3%), “Ukraine’s International Image in the EU” (2/15, 13.3%), “Law Enforcement Agencies of Ukraine” (2/15, 13.3%), and “Armed Forces of Ukraine” (2/15, 13.3%).

Thus, the results of the 15 analyzed cases demonstrated that:

- The sentiment percentage ranged from −10% to +30%, with the following distribution:
    a. Positive (≥ +10%) — 8 cases (53.3%);
    b. Negative (≤ −10%) — 4 cases (26.7%);
    c. Neutral (0%) — 3 cases (20%).

Table 2
Comparative analysis of 15 case studies

№	Source	Category	Sentiment	Article of the Law / Code
1	BBC News, At least eight people killed and more than 80 injured in overnight attack on Kyiv, BBC News, April 24, 2025 [28]	Ukraine's International Image in the USA, Canada, and the United Kingdom	+10 %	Not entered
2	Hromadske, Тіло журналістки Рощиної, яку закатували в полоні, повернули в Україну, Hromadske, 2025 [29]	Armed Forces of Ukraine	+10 %	Not entered
3	Hromadske Rehiony, На Тернопільщині почали ексгумацію польських жертв Волинської трагедії, Hromadske, 2025 [30]	Military-political leadership of Ukraine at all levels	+20 %	Not entered
4	BBC News, Pope Francis to be buried at Santa Maria Maggiore, BBC News, April 24, 2025 [31]	The international image of Ukraine in the EU (in 8 languages)	−10 %	Not entered
5	Hromadske, Вижити в «автобусі смерті». Репортаж із Сум, Hromadske, 2025 [32]	Military-political leadership of Ukraine at all levels	+10 %	Not entered
6	Fox News, Key Karen Read witness admits grand jury testimony wasn't true, Fox News, 2025 [33]	Armed Forces of Ukraine	−10 %	Not entered
7	Magnolia-TV, У Києві чоловіка судитимуть за обвинуваченням у вуличному збуті психотропів, Magnolia-TV, 2025 [34]	Law enforcement agencies of Ukraine	−10 %	Constitution of Ukraine, Part 4, Article 32
8	Bessarabia Inform, Мешканець Болградського району отримав умовний термін за незаконне зберігання зброї та наркотиків, Bessarabia Inform, April 2025 [35]	Law enforcement agencies of Ukraine	0 %	Not entered
9	UA.News, Стрілянина в Подільському районі Києва: поліція з'ясовує обставини інциденту, UA.News, 2025 [36]	Military-political leadership of Ukraine at all levels	+10 %	Criminal Code of Ukraine, Article 109
10	LIGA.net, Що означає скасування наказу про призов Насірова до ЗСУ?, LIGA.net, 2025 [37]	Military-political leadership of Ukraine at all levels	0 %	Civil Code of Ukraine, Article 278
11	0462.ua, НАБУ закрило справу про незаконне збагачення нардепа Павла Халімона, 0462.ua, 2025 [38]	Military-political leadership of Ukraine at all levels	+10 %	Civil Code of Ukraine, Article 278
12	Interfax-Ukraine, П'ять діб ліквідували пожежу у Балаклійському лісництві, Interfax-Ukraine, 2025 [39]	Military-political leadership of Ukraine at all levels	0 %	Constitution of Ukraine, Part 4, Article 32
13	2day.kh.ua, На Харківщині загасили масштабну лісову	Military-political leadership of Ukraine at all levels	+30 %	Criminal Code of Ukraine, Article 259

	пожежу, що вирувала п'ять діб, 2day.kh.ua, 2025 [40]			
14	Харків Times, Рятувальники Харківщини ліквідували пожежу в Балаклійському лісництві, Харків Times, April 24, 2025 [41]	The international image of Ukraine in the EU (in 8 languages)	+30 %	Criminal Code of Ukraine, Article 182
15	Life Kyiv UA, Стрілянина серед білого дня на Подолі: що сталося і хто стріляв, Life Kyiv UA, 2025 [42]	Military-political leadership of Ukraine at all levels	−10 %	Civil Code of Ukraine, Article 278

- The average sentiment value was +6%, the median was +10%, and the standard deviation was approximately 11.5%..
- In 5 out of 15 cases (33.3%), a relevant legal article was automatically identified:
    a. Constitution of Ukraine, Part 4 Article 32 (2 cases);
    b. Civil Code of Ukraine, Article 278 (2 cases);
    c. Criminal Code of Ukraine, Article 109 (1 case);
    d. Criminal Code of Ukraine, Article 182 (1 case);
    e. Criminal Code of Ukraine, Article 259 (1 case).
- In 10 cases (66,7 %) the keyword occurrence threshold (τ=5) was not reached, and therefore no relevant article was identified.

The highest percentage of successful matches to legal articles was observed in the categories "Law Enforcement Agencies of Ukraine" (1/2, 50%) and "Military-Political Leadership of Ukraine" (3/7, 42.9%). In the thematic "international image" categories, no correct matches were identified (0/4, 0%), indicating insufficient representation of relevant key word forms in these groups.

The obtained quantitative indicators suggest satisfactory accuracy in sentiment analysis of texts, but a limited ability of the algorithm to identify specific legal articles.

To improve the frequency of successful matches to legal norms, expert dictionaries should be expanded, enriched with synonymous forms, and consideration should be given to lowering the threshold τ for shorter news materials.

5. Conclusions

The conducted evaluation showed that the proposed rule-based method for automatically aligning media texts with articles of Ukrainian legislation demonstrates an overall accuracy of 83.8%, corresponding to Precision = 0.84, Recall = 0.83, and F1-score = 0.84 on a control sample of 12 legal classes. These metrics are only slightly lower than the results of previous rule-based studies in medical Q&A domains (Accuracy > 85%) and social research (Accuracy = 88%) [16, 18], indicating the competitiveness of the approach even within the highly specialized legal domain. At the same time, our metrics demonstrate a high balance between precision and recall, which is critical for real-world media monitoring tasks where classification errors may have significant legal implications.

A detailed analysis of the confusion matrix revealed that most errors occurred between closely related categories such as "Military-Political Leadership of Ukraine" and "Law Enforcement Agencies of Ukraine" (K1↔K2, K6↔K7), which share several key terms. Furthermore, in a series of 15 real-world case studies, the algorithm successfully identified a relevant article in only 33.3% of cases (5/15), which is likely due to the fixed threshold τ = 5 and the relatively short length of typical news content. The highest rate of successful matches was recorded in the category "Law Enforcement Agencies of Ukraine" (50%), and the lowest — in the international image categories (0%), highlighting the need to further refine expert dictionaries specifically for "international" topics.

The obtained results highlight the value of the proposed Explainable AI tool: the system operates without the need for trained models, offering full transparency of its decision logic and flexible terminology customization through the user interface. This architecture is particularly beneficial for governmental institutions and legal firms, where it is critically important to reproduce why and according to which rules a text was matched to a specific legal article. Furthermore, the combination of normative alignment with sentiment analysis opens opportunities for comprehensive monitoring of reputational risks and compliance with media standards.

Future research will focus on expanding and enriching expert dictionaries with new synonyms, idiomatic expressions, and polysemous lexemes; adapting the threshold τ based on text length and genre to increase sensitivity to shorter messages; integrating semantic methods (e.g., word embeddings or topic ontologies) to reduce inter-class confusion; developing hybrid approaches that combine rule-based logic with lightweight statistical or transformer modules; and enabling multilingual support and cross-national comparative analysis of legal texts. Together, these enhancements aim to significantly improve the system's legal coverage, increase the frequency of successful matches, and broaden its practical applications in legal analysis of media content.

Declaration on Generative AI

The authors used GPT-4 and DeepL to prepare this paper: Grammar and Spelling Checker. After using these tools, the authors reviewed and edited the content as necessary and are solely responsible for the content of the publication.

References

[1] A. Sachenko, T. Lendiuk, K. Lipianina-Honcharenko, M. Dobrowolski, G. Boguta, and L. Bytsyura, "Method of determining the text sentiment by thematic rubrics", in *Intell. Syst. Workshop CoLInS 2024*. CoLInS, 2024. Accessed: Apr. 24, 2025. [Online]. Available: https://doi.org/10.31110/colins/2024-3/026

[2] M. Komar, V. Dorosh, G. Hladiy, A. Sachenko, Deep neural network for detection of cyber attacks, in: Proceedings of the 2018 IEEE 1st International Conference on System Analysis and Intelligent Computing (SAIC 2018), IEEE, 2018. Article ID: 8516753.

[3] M. Komar, A. Sachenko, S. Bezobrazov, V. Golovko, Intelligent cyber defense system using artificial neural network and immune system techniques, Communications in Computer and Information Science 783 (2017) 36–55.

[4] M. Dyvak, A. Melnyk, Y. Kedrin, Interval model of the user reactions to messages in thematic groups of social networks, in: 2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), IEEE, 2022, pp. 837–840.

[5] O. Kovalchuk, M. Kasianchuk, M. Karpinski, R. Shevchuk, Decision-making supporting models concerning the internal security of the state, International Journal of Electronics and Telecommunications (2023) 301–307.

[6] M. Raees, S. Fazilat, *Lexicon-based sentiment analysis on text polarities with evaluation of classification models*, arXiv preprint arXiv:2409.12840, 2024.

[7] D. H. Abd, A. R. Abbas, A. T. Sadiq, Analyzing sentiment system to specify polarity by lexicon-based, *Bulletin of Electrical Engineering and Informatics* (2021). URL: https://beei.org/index.php/EEI/article/view/2471.

[8] A. Ullah, S. N. Khan, N. M. Nawi, Review on sentiment analysis for text classification techniques, *Multimedia Tools and Applications* (2023). URL: https://link.springer.com/article/10.1007/s11042-022-14112-3.

[9] S. V. Balshetwar, R. M. Tuganayat, Frame tone and sentiment analysis, *ResearchGate preprint* (2019). URL: https://www.researchgate.net/publication/335811449_Frame_Tone_and_Sentiment_Analysis.

[10] R. Catelli, S. Pelosi, C. Comito, C. Pizzuti, Lexicon-based sentiment analysis on Twitter in Italy, *Computers in Biology and Medicine* (2023). URL: https://www.sciencedirect.com/science/article/pii/S0010482523003414.

[11] K. K. Kiilu, Sentiment Classification for Hate Tweet Detection in Kenya, *Master's thesis, Jomo Kenyatta University of Agriculture and Technology* (2021). URL: http://ir.jkuat.ac.ke/handle/123456789/5521.

[12] R. Satapathy, C. Guerreiro, I. Chaturvedi, Phonetic-based microtext normalization, in: *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, 2017. URL: https://ieeexplore.ieee.org/abstract/document/8215691/.
[13] M. Itani, Sentiment analysis for informal Arabic text, *Ph.D. thesis, Sheffield Hallam University* (2018). URL: https://shura.shu.ac.uk/23402/1/Itani_2018_phd_SentimentAnalysisAnd.pdf.
[14] R. Guarasci, R. Catelli, M. Esposito, Deceptive reviews detection in cultural heritage domain, *Expert Systems with Applications* (2024). URL: https://www.sciencedirect.com/science/article/pii/S0957417424009977.
[15] M. Kumar, L. Khan, H. T. Chang, Evolving techniques in sentiment analysis, *PeerJ Computer Science* (2025). URL: https://peerj.com/articles/cs-2592/.
[16] Y. Zhang, T. Wang, Y. Wang, J. Cao, *Knowledge discovery of diseases symptoms and rehabilitation measures in Q&A communities*, 2025. URL: https://www.nature.com/articles/s41598-025-98300-9.
[17] B. Lashkari, *Classification and Vulnerability Detection of Ethereum Energy Smart Contracts*, 2024. URL: https://era.library.ualberta.ca/items/5d6a38c4-f7f8-4c24-bb5c-23ba247ac1e7.
[18] B. E. Perron, H. Luan, B. G. Victor, *Moving Beyond ChatGPT: Local Large Language Models (LLMs) and the Secure Analysis of Confidential Unstructured Text Data in Social Work Research*, 2024. URL: https://journals.sagepub.com/doi/abs/10.1177/10497315241280686.
[19] A. Thöni, *Sustainability risk monitoring in supply chains: ranking suppliers using text mining and Bayesian networks with a focus on child labor*, Dissertation, Technische Universität Wien, 2015. URL: https://doi.org/10.34726/hss.2015.30640.
[20] W. J. Yeo, W. Van Der Heever, R. Mao, E. Cambria, *A comprehensive review on financial explainable AI*, 2025. URL: https://link.springer.com/article/10.1007/s10462-024-11077-7.
[21] S. Narayanan, P. G. Georgiou, *Behavioral signal processing: Deriving human behavioral informatics from speech and language*, *Proc. IEEE* 101 (5) (2013) 1203–1233.
[22] F. Chiarello, *Mining Technical Knowledge*, Ph.D. thesis, 2019. URL: https://tesidottorato.depositolegale.it/handle/20.500.14242/131397.
[23] R. Nai, *Analysis of Public Administration Procurement and Expenditures Related to Energy Efficiency Improvements*, Ph.D. thesis, 2025. URL: https://tesidottorato.depositolegale.it/handle/20.500.14242/199436.
[24] X. Liu, C. Li, *Artificial intelligence and translation*, in: *Routledge Encyclopedia of Translation Technology*, Routledge, 2023, pp. 280–302.
[25] K. Lipianina-Honcharenko, A. Melnychuk, K. Yurkiv, G. Hladiy, M. Telka, *Integrated Approach to the International Aspects of Online Dispute Resolution Formation*, in: *Proceedings of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development*, Ternopil, Ukraine, May 2024, pp. 88–98.
[26] K. Lipianina-Honcharenko, D. Lendiuk, M. Nazar Melnyk, T. L. Komar, *Evaluation of the Keyword Selection Methods Effectiveness for the Fake News Classification*, 2024.
[27] *Streamlit app*, Streamlit, n.d. Retrieved April 24, 2025. URL: https://nelczgkwcsghtwdkw7buxn.streamlit.app/.
[28] BBC News, *At least eight people killed and more than 80 injured in overnight attack on Kyiv*, BBC News, April 24, 2025. URL: https://www.bbc.com/news/articles/cd7v0lgg18xo.
[29] Hromadske, *Тіло журналістки Рощиної, яку закатували в полоні, повернули в Україну*, Hromadske, 2025. URL: https://hromadske.ua/viyna/243651-tilo-zurnalistky-roshchynoyi-iaku-zakatuvaly-v-poloni-povernuly-v-ukrayinu-zastupnyk-hlavy-mzs.
[30] Hromadske Rehiony, *На Тернопільщині почали ексгумацію польських жертв Волинської трагедії*, Hromadske, 2025. URL: https://hromadske.ua/rehiony/243654-na-ternopilshchyni-pochaly-ekshumatsiiu-polskykh-zertv-volynskoyi-trahediyi-rmf-fm.
[31] BBC News, *Pope Francis to be buried at Santa Maria Maggiore*, BBC News, April 24, 2025. URL: https://www.bbc.com/news/articles/cgrgzl0vyqpo.
[32] Hromadske, *Вижити в «автобусі смерті». Репортаж із Сум*, Hromadske, 2025. URL: https://hromadske.ua/viyna/243172-vyzyty-v-avtobusi-smerti-reportaz-iz-sum.
[33] Fox News, *Key Karen Read witness admits grand jury testimony wasn't true*, Fox News, 2025. URL: https://www.foxnews.com/us/key-karen-read-witness-admits-grand-jury-testimony-wasnt-true.
[34] Magnolia-TV, *У Києві чоловіка судитимуть за обвинуваченням у вуличному збуті психотропів*, Magnolia-TV, 2025. URL: https://magnolia-tv.com/news/124548-u-kyyevi-

cholovika-sudytymut-za-obvynuvachennyam-u-vulychnomu-zbuti-psykhotropiv?prov=ukrnet.
[35] Bessarabia Inform, *Мешканець Болградського району отримав умовний термін за незаконне зберігання зброї та наркотиків*, Bessarabia Inform, April 2025. URL: https://bessarabiainform.com/2025/04/meshkanets-bolhradskoho-rayonu-otrymav-umovnyy-termin-za-nezakonne-zberihanni-zbroi-ta-narkotykiv/.
[36] UA.News, *Стрілянина в Подільському районі Києва: поліція з'ясовує обставини інциденту*, UA.News, 2025. URL: https://ua.news/ua/capital/strilianina-v-podils-komu-raioni-kiieva-politsiia-ziasovuie-obstavini-intsidentu.
[37] LIGA.net, *Що означає скасування наказу про призов Насірова до ЗСУ?*, LIGA.net, 2025. URL: https://news.liga.net/ua/politics/news/shcho-oznachaie-skasuvannia-nakazu-pro-pryzov-nasirova-do-zsu-liganet-otrymala-poiasnennia.
[38] 0462.ua, *НАБУ закрило справу про незаконне збагачення нардепа Павла Халімона*, 0462.ua, 2025. URL: https://www.0462.ua/news/3935917/nabu-zakrilo-spravu-pro-nezakonne-zbagacenna-nardepa-vid-slugi-narodu-pavla-halimona-z-cernigivsini.
[39] Interfax-Ukraine, *П'ять діб ліквідували пожежу у Балаклійському лісництві*, Interfax-Ukraine, 2025. URL: https://interfax.com.ua/news/general/1066345.html.
[40] 2day.kh.ua, *На Харківщині загасили масштабну лісову пожежу, що вирувала п'ять діб*, 2day.kh.ua, 2025. URL: https://2day.kh.ua/ua/kharkow/na-kharkivshchyni-zahasyly-masshtabnu-lisovu-pozhezhu-shcho-vyruvala-pyat-dib.
[41] Харків Times, *Рятувальники Харківщини ліквідували пожежу в Балаклійському лісництві*, Харків Times, April 24, 2025. URL: https://times.kharkiv.ua/2025/04/24/ryatuvalniki-harkivshhini-likviduvali-pozhezhu-v-balaklijskomu-lisnitstvi/.
[42] Life Kyiv UA, *Стрілянина серед білого дня на Подолі: що сталося і хто стріляв*, Life Kyiv UA, 2025. URL: https://life.kyiv.ua/news/strilyanina-sered-bilogo-dnya-na-podoli-scho-stalosya-i-hto-strilyav.