

# Evaluation of methods for intellectual analysis of manipulative content in the information space\*

Vsevolod Senkivskyy<sup>1,†</sup>, Iryna Pikh<sup>1,†</sup>, Alona Kudriashova<sup>1,\*,†</sup>, Roman Andriiv<sup>2,†</sup> and Nazarii Senkivskyy<sup>1,†</sup>

<sup>1</sup> Lviv Polytechnic National University, Stepan Bandera Str., 12, Lviv, 79013, Ukraine

<sup>2</sup> Separate structural unit "Berezhany Professional College of NUBiP of Ukraine"

20 Akademichna St., Berezhany, Ternopil region, Ukraine, 47501

## Abstract

The modern information space is characterized by a high level of manipulative content, which poses a serious threat and significantly affects public opinion, political processes and the level of trust in media resources. Determining the optimal methods for evaluating the veracity of information messages is an extremely urgent task that requires the use of intelligent data analysis and efficient means of counteraction. The attention is focused on the use of machine learning algorithms and artificial intelligence to detect manipulations in text, video and audio materials.

Based on the developed expert survey methodology, logistic regression, decision trees and SVM (Support Vector Machine) are identified among the known methods of intelligent analysis of manipulative content. To determine the most efficient approach, a Pareto set of mutually non-dominated criteria is formed, which includes accuracy, execution speed, resource capacity and interpretability.

A model of the priority influence of criteria on the message evaluation process is developed, which allows determining their weight coefficients as the basis for experimental calculations. In order to select the optimal method, the method of linear convolution of criteria is applied, which ensures the determination of the Pareto-optimal solution among the proposed alternatives.

To implement this approach, alternative options are generated by varying the criteria and forming combinations of their efficiency degrees that correspond to the methods of evaluating the message probability. Based on the method of hierarchy analysis, a program calculation of normalized weights of criteria and utility functions for the corresponding alternatives is performed. The obtained final values of the combined functionalities of alternative options allow determining the optimal method for evaluating the probability of information messages.

The results of the study demonstrate the dependency of the method selection on the requirements for accuracy and explainability of solutions. The obtained data can be used to develop automated systems for monitoring and analysing the information space, which will contribute to increasing the level of information security and efficient counteraction to manipulative content.

## Keywords

manipulative content, information message, intelligence analysis, machine learning, information security, method of linear convolution of criteria, alternative option, pairwise comparison matrix, utility function, Pareto-optimal method

## 1. Introduction

In the context of growing information threats, there is a growing need to develop and improve methods of intelligent analysis that allow for the efficient detection, identification and evaluation of false content. The research in this area is aimed at developing algorithmic and analytical approaches that provide the detection of signs of manipulation, analysis of information sources and determination of its reliability level [1, 2]. The use of modern data analysis technologies, in particular

---

*Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security, April 04, 2025, Khmelnytskyi, Ukraine*

\* Corresponding author.

† These authors contributed equally.

✉ vsevolod.m.senkivskyy@lpnu.ua (V. Senkivskyy); iryna.v.pikh@lpnu.ua (I. Pikh); alona.v.kudriashova@lpnu.ua (A. Kudriashova); roman.andriiv@icloud.com (R. Andriiv); nazarii.y.senkivskyy@lpnu.ua (N. Senkivskyy)

ORCID: 0000-0002-4510-540X (V. Senkivskyy); 0000-0002-9909-8444 (I. Pikh); 0000-0002-0496-1381 (A. Kudriashova); 0009-0001-2450-5439 (R. Andriiv); 0009-0009-0467-2828 (N. Senkivskyy)



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

machine learning and artificial intelligence methods, increases the accuracy of information flow evaluation and prevents the negative impact of destructive content on the society [3]. The development of artificial intelligence, machine learning, linguistic analysis and neural networks contributes to the improvement of methods for processing text and multimedia data [4, 5]. The publication [6] analyzes the efficiency of artificial intelligence and machine learning methods in cybersecurity, evaluating their ability to detect threats and anomalies. Various approaches, their advantages, disadvantages and development prospects are considered. A feature of the issues studied in this paper is the coverage of approaches in the work [7], which combine computational solutions and propose strategies to combat disinformation, in particular, the use of machine learning-based methods for automatic classification of disinformation. Machine learning methods use statistical approaches [8, 9] to identify patterns and anomalous deviations in large data sets, which allows security analysts to timely detect potential threats, including previously unknown attacks. Deep learning, as a subfield of machine learning, demonstrates high efficiency in increasing the accuracy of recognizing malicious patterns due to multi-level information processing, which is especially important for the analysis of complex cyber threats, including image recognition, traffic analysis, and decryption of threatening messages.

The efficiency evaluation of such methods can be based on the Pareto principle, which allows selecting optimal solutions based on the criteria of accuracy, speed, and resource consumption. For example, deep learning models [10, 11] demonstrate high accuracy, but require significant computing resources, while classical machine learning methods are less expensive, but may be inferior in the ability to recognize complex manipulative structures. The selection of a specific approach depends on the characteristics of the information environment in which the analysis is carried out, as well as on the needs for accuracy and speed of data processing [12].

In view of the above, the goal of the study is to evaluate the efficiency of modern methods for intelligent analysis of manipulative content by comparative analysis of their accuracy, speed, resource capacity, and interpretability. Special attention is paid to the criteria for selecting the optimal method, which allows increasing the efficiency of information security and minimizing the negative impact of disinformation on the society.

Strategies for evaluating methods of intelligent analysis of manipulative content based on multi-criteria optimization are proposed. This work analyzes modern methods of detecting manipulative content, identifies key criteria for their evaluation, ranks alternatives using the method of linear convolution of criteria and determines the optimal method. The results obtained can be used to develop automated systems for monitoring the information space, which will help reduce the risks of spreading disinformation and increase the level of information security.

## **2. Literature review**

Research and analysis of manipulative content in the information space covers a wide range of methods and approaches aimed at identifying, assessing and neutralizing the impact of false information. This includes improving fake detection algorithms, developing multi-criteria analysis systems, and integrating modern machine learning technologies to improve the analysis accuracy. The use of efficient methods is critical for protecting various areas of activity exposed to the risks of disinformation and manipulation.

Although most research focuses on detecting fake news [13] based on its content or using the user interaction with news on social networks, there is a growing interest in proactive intervention strategies to counter the spread of disinformation and its impact on society [14]. However, the selection of the optimal method remains an open question, since the efficiency of algorithms significantly depends on the characteristics of the input data, the context of their application, the level of adaptability to new manipulative strategies, and the performance of computing resources.

In the article [15], multimodal methods for detecting fake news based on semantic information have achieved great success. However, these methods use only the deep features of multimodal information, which leads to a large loss of real information at the surface level.

For a well-grounded selection of the optimal method, a set of interrelated criteria is taken into account by the authors. In this context, the use of multi-criteria analysis methods plays a special role, such as the method of linear convolution of criteria, which allows evaluating alternative approaches by key parameters. The formation of the Pareto set allows one to isolate mutually non-dominated methods and determine the best option based on the calculation of utility functions.

The study [16] presents a taxonomy of models, machine learning and deep learning functions, used to detect fake news based on the content analysis. To solve this problem, machine learning models are used that allow for automated identification of unreliable information. At the same time, the work did not propose efficient methods for improving the accuracy of classification or optimizing models, which leaves open the questions of their adaptation to new types of manipulative content and increasing resistance to changes in the characteristics of fake news.

The use of neural network technologies for detecting fake data is considered in the work [17]. The efficiency of the methods is assessed by comparative analysis of their strategies, approaches to error estimation and the accuracy level on different data sets. Our goal is to help researchers in determining relevant criteria and selecting the optimal method for solving specific tasks of intelligent analysis of manipulative content.

A detailed review of methods for generating and detecting deep fakes is presented in the study [18]. Open challenges faced by detection systems are considered, and possible options for overcoming them are proposed, in particular using deep learning. The main emphasis is placed on the need to create efficient manipulative content detection systems that are able to adapt to new challenges in the field of artificial intelligence. However, there are no experimental results and practical evaluation of the proposed methods in real-world application scenarios. The work [19] is devoted to the analysis of the problem of large language model radicalization. Semantic vulnerabilities and the learning inadequacy based on human feedback are studied. At the same time, the attention is focused mainly on theoretical aspects and does not contain practical recommendations for the direct implementation of protective mechanisms.

The work [20] is focused on the problem of manipulation by artificial intelligence and recommends criteria for assessing the level of the system manipulability. The main attention is paid to the analysis of algorithmic ethics and the need to create transparent decision-making systems that are not susceptible to manipulation by users. However, there are no clear metrics or practical methods for assessing the manipulability level.

To solve this problem, improved approaches to fake news detection are proposed by the authors, which combine traditional machine learning methods with the optimal selection of data pre-processing. It is assumed to use the method of linear convolution of criteria to integrate various performance indicators of models related to manipulative content. The proposed approach will contribute to increasing the reliability of classification and the balance between accuracy and computational efficiency. Summarizing, the research highlights a wide range of threats related to manipulative content and artificial intelligence. This emphasizes the importance of a comprehensive approach to countering such threats, which involves improving detection methods, applying multi-criteria analysis, and integrating modern machine learning technologies.

### **3. Material and methods**

#### **3.1. Pareto principle and selection of a criteria set**

The Pareto principle, or 80/20 rule, is a fundamental approach to analysing efficiency and optimality in various fields, including multi-criteria analysis and decision-making. It states that in many processes, approximately 80% of the results are due to 20% of the factors. In the context of optimization and efficiency analysis, this principle is used to identify the most important parameters that have the greatest impact on the final result [21, 22].

In multi-criteria analysis, the Pareto principle is applied through the concept of Pareto optimality, which defines a set of solutions that cannot be improved on one criterion without

worsening another. This means that no solution in the Pareto set is dominant over another, and the selection of a particular option depends on priorities or additional conditions set by the researcher or the receiving party.

The formation of a Pareto set of criteria involves analysing the set of possible solutions and selecting those that are Pareto-efficient. If a solution is not inferior to others in any criterion and has an advantage in at least one, it is considered Pareto-optimal and is included in the corresponding set. In the case of a large number of criteria and a significant selection of solutions, evolutionary search methods or stochastic approaches can be used.

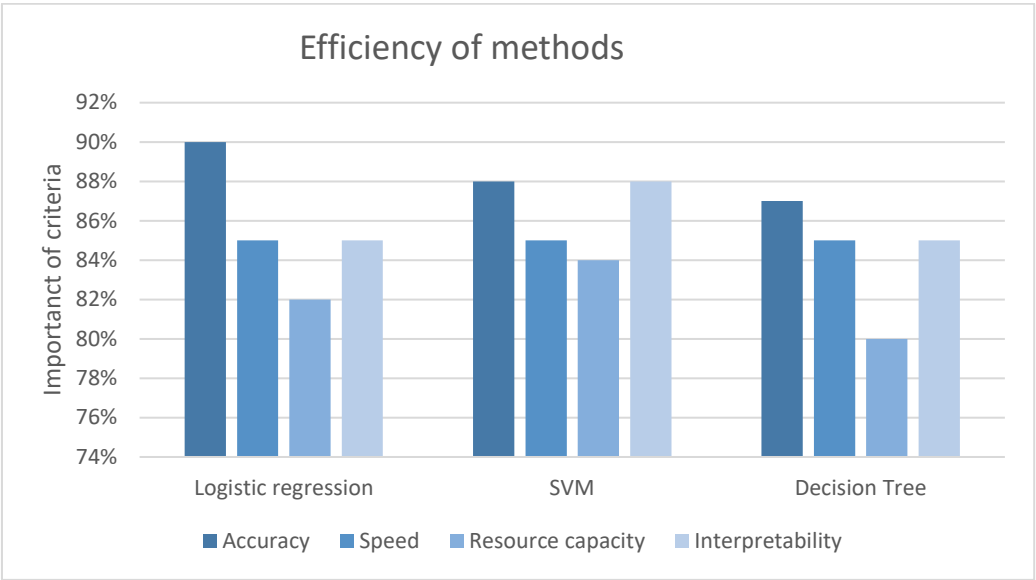
The criteria influencing the process of determining the optimal method for detecting manipulative content are presented in Table 1.

**Table 1**  
Pareto-set of criteria for alternative methods

Methods	Criteria for the efficiency of alternative methods			
	Accuracy	Performance Speed	Resource capacity	Interpretability
Logistic regression	90%	85%	82%	85%
SVM (support vector machine)	88%	85%	84%	88%
Decision Tree	87%	85%	80%	85%

A feature of the Pareto set is its adaptability to a specific task: it allows one to identify compromise solutions that take into account different aspects of the problem, and provides a basis for further analysis and decision-making. Since none of the options is unambiguously the best according to all criteria, the selection of the final solution is often made taking into account additional priorities or weight factors that represent the significance of each criterion in a specific context.

Figure 1 presents a diagram of the efficiency of the criteria assigned to the methods of detecting manipulative content in the information space.



**Figure 1:** Diagram of the criteria efficiency for the methods of message probability evaluation.

Thus, the use of the Pareto principle in multi-criteria intelligent analysis of manipulative content allows for efficient resource allocation, finding optimal solutions, and minimizing trade-offs between conflicting requirements. This is especially relevant in complex text message processing systems, where it is necessary to obtain the interaction of many factors and their influence on the final result.

### 3.2. Formation of an expert group

The use of expert reviews in the evaluation of methods for intellectual analysis of manipulative content allows one to obtain a quantitative assessment of the importance degree of each of the criteria that form a set of values of factors influencing the process quality [22]. In this case, the scale rating method is used, which provides for obtaining quantitative assessments of the importance degree of each of the factors belonging to a certain set, relative to the scale of their basic (reference) values. In this case, the estimates of the relative importance of each factor are expressed in points on a certain scale. The most commonly used is a 100-point scale, where the maximum possible importance corresponds to a score of 100 points, and the minimum possible one is 0 (zero) points.

When processing expert data, the survey results are summarized in a table, where  $C_{ji}$  is the relative importance of the parameter  $x_i$  from the point of view of the  $j$ -th expert, which is expressed by the corresponding score or rank value (Table 2).

**Table 2**

Results of experts' survey using the scale rating method

Expert	Factor (parameter)					
	$x_1$	$x_2$	...	$x_i$	...	$x_n$
1	$C_{11}$	$C_{12}$	...	$C_{1i}$	...	$C_{1n}$
2	$C_{21}$	$C_{22}$	...	$C_{2i}$	...	$C_{2n}$
...	...	...	...	...	...	...
$j$	$C_{j1}$	$C_{j2}$	...	$C_{ji}$	...	$C_{jn}$
...	...	...	...	...	...	...
$m$	$C_{m1}$	$C_{m2}$	...	$C_{mi}$	...	$C_{mn}$

The arithmetic mean of factor scores is determined from the expression:

$$C_i = \frac{1}{m} \sum_{j=1}^{m_i} C_{ji} \quad (1)$$

where  $m_i$  is a number of experts evaluating the importance of factors  $x_i$ .

The values  $C_{ji}$  and  $C_i$  can be expressed quantitatively in points or ranks. In the first case, the value is called the average score (average value) of the criterion  $x_i$ , in the second case it is the average rank. Summing the numbers  $i, j$  in the rows with the subsequent division of the obtained result by  $m$  gives the average ranking of the factors  $x_1, x_2, \dots, x_n$ , which, in turn, serves as an indicator of the generalized opinion about the importance of the factors (the smaller the sum in the row  $j$  is, the more important role the factor  $i$  plays). The opposite picture occurs with respect to the sums in the columns.

Since the indicator of the generalized opinion  $C_i$  and the reference value essentially differ only in their purpose, in the following, for the sake of simplicity of reasoning, one will consider the centre of grouping of scale scores, considering that this concept covers the previous two.

The method of searching the centre of grouping of expert data on the rating scale for any distribution law uses the average, or weighted average score value. This approach (especially the use of the weighted average value) allows to objectively determine the centre of grouping with a sufficient degree of approximation [20]. However, with a large range of scale values, taking into account all values without exception, as it will be shown below, can lead to a significant shift in the centre of grouping.

Suppose the centre of score grouping for a given distribution of experts by the values provided by them is denoted by  $C$ . The following relation is valid:

$$C = F(k, h, W), \quad (2)$$

where:  $k$  is a number of experts in the group;  $h$  is a step for searching a grouping area;  $W$  is a range of score values, to which the number of experts corresponds, is not less than  $\theta_k$ , at the step ( $0 < \theta_k < 1$ ).

Suppose one has a scale with values  $i$ , where  $i = 0, 1, 2, \dots, n$ . Then  $m_i$  is a number of experts who put the  $i$ -th value. If the group includes  $k$  experts, then  $\sum_{i=1}^n m_i = k$ . In the first step of the search ( $h = 1$ ) of the center of grouping, pairs of values are defined that satisfy the expression:

$$\sum_{i \in W_h}^n m_i \geq \theta_k \quad (3)$$

In this case, the following three logical options are possible:

- no pair of values at this step satisfies the relation (2). Then the searching step of the grouping area increases by one, that is, the "weighting" area expands, and the procedure is repeated;
- there is exactly one area on the scale at this searching step that satisfies the relation (3). In this case, the grouping area is identified and the centre of grouping is determined as the weighted average of all values belonging to this area:

$$C = \frac{\sum_{i \in W_h}^n i \cdot m_i}{\sum_{i \in W_h}^n m_i} \quad (4)$$

- there are several areas satisfying the relation (4). Then the grouping area is defined as follows: the left boundary represents the smallest value for all values of the these areas, and the right boundary represents their largest value, respectively. The grouping area is identified as the weighted average of all values belonging to the grouping area:

$$C = \frac{\sum_{i \in G}^n i \cdot m_i}{\sum_{i \in G}^n m_i} \quad (5)$$

where  $G$  is a set of all scale values belonging to the grouping area.

Suppose the total set of criteria for evaluating the process efficiency is 8. Using expert evaluation, a subset of four factors should be selected (Table 3). To solve the problem, a scale from 0 to 8 and the number of experts corresponding to each scale value  $m_i = 50$  are used.

**Table 3**  
Formation of expert groups

$i$	1	2	3	4	5	6	7	8	9	10	11	12
$m_i$	8	6	7	3	4	7	4	5	1	2	2	1

Suppose one takes  $\theta = 0,5$ , since the 50% comparison is the most common in the expert evaluation method. For the total number of experts  $m_i = 50$ , the search step is set  $h = 1$  and the number of experts for each area is calculated (Table 4).

**Table 4**  
Quantitative composition of experts in each area during the first step

$W_1$	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
$\sum_{i \in W_1} m_i$	14	13	10	7	11	11	9	6	3	4	3

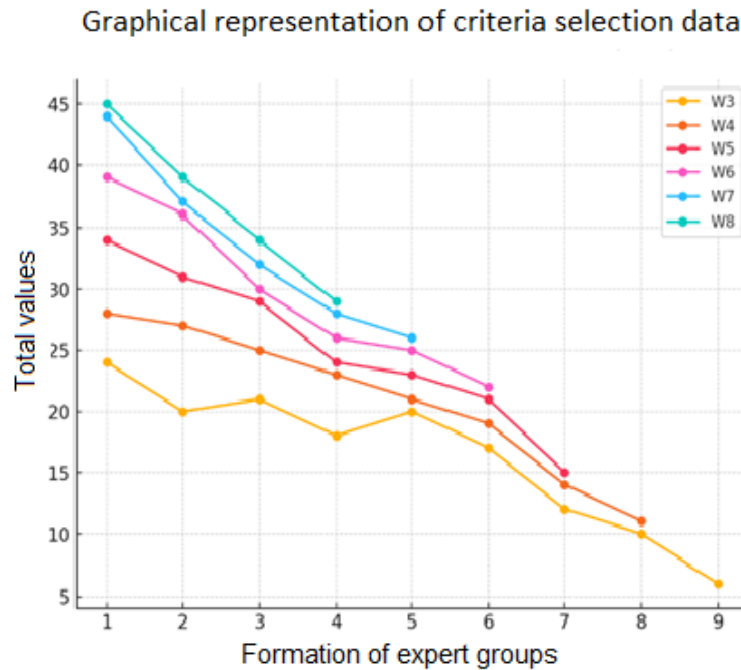
According to the last table, no area satisfies the relation (2) at a given step. Therefore, the searching step is increased ( $h = 2$ ) and the expanded areas are calculated (Table 5).

**Table 5**

Quantitative composition of experts in each area during the second step

$W_2$	1-3	2-4	3-5	4-6	5-7	6-8	7-9	8-10	9-11	10-12
$\sum_{i \in W_2} m_i$	21	16	14	14	15	16	10	8	5	5

As it can be seen from the data presented, at this step there is also no area satisfying the relation (3). Therefore, the searching step is increased again and the procedure is repeated until four optimal influencing criteria are determined (Fig. 2).



**Figure 2:** Diagram of determining optimal influencing criteria.

As it follows from the calculations, there is exactly one area that satisfies the relation (4), namely area 1-8. It becomes the grouping area.

Using the relation (5), the value of the centre of grouping is calculated:

$$C = \frac{1 \times 8 + 2 \times 6 + 3 \times 7 + 4 \times 3 + 5 \times 4 + 6 \times 7 + 7 \times 4 + 8 \times 5}{8 + 6 + 7 + 3 + 4 + 7 + 4 + 5} = \frac{183}{44} = 4$$

With a weighted average value of the entire scale  $C = \frac{246}{50} = 5$ .

Thus, the method of generalizing scale scores during group examination ensures the formation of the most reliable indicator of generalized opinion for any distribution of expert data – the centre of grouping of scores, which, with a symmetric distribution of scores, coincides with the weighted average value.

### 3.3. Method of linear convolution of criteria

The multilevel models of the priority influence of technological indicators on the level of performance of publishing and printing processes obtained in the work [24] serve as the basis for making a decision on the importance of a particular factor and its “contribution” to the overall quality

of the resulting product. Using mathematical terminology, one has the necessary initial data on the degree of the indicator influence on the process, but they are insufficient for their full practical implementation. Knowledge is essential not only about the conditional importance of the technological factor. It is necessary to study how much effort should be spent on each of the indicators during their interaction to achieve proper process efficiency.

The issues outlined in this paper concern the generation of alternative options based on methods for processing text information messages and selecting the optimal solution in accordance with the specified principle. This task belongs to the field of multi-criteria optimization, which involves making a reasoned decision on selecting the most efficient method for evaluating the message reliability.

To solve this problem, it is necessary to determine a set of criteria, the composition and content of which are formed on the basis of the Pareto principle [25]. The main idea of the approach is that the criteria included in the set of Pareto-optimal solutions cannot be completely dominated by other criteria. That is, there is no criterion that would surpass all the others in all indicators at the same time. Within the framework of the study, based on expert evaluation, mutually non-dominated criteria are selected that form the Pareto set, namely: accuracy, execution speed, resource capacity and interpretability.

According to the methods of decision theory [26, 27], multi-criteria optimization on a set of alternatives  $D$  in the presence of objective functions  $f(x) = (f_1(x), \dots, f_m(x))$  involves constructing models of utility functions and determining their maximum value:  $f_i(x) \rightarrow \max_{x \in D} f_i(x), i = 1, m$ .

The process of multi-criteria selection of the optimal alternative is based on the method of linear convolution of criteria, which involves linear combination of partial target functionals  $f_1, \dots, f_m$  into a single generalized functional:

$$F(w, x) = \sum_{i=1}^m w_i f_i(x) \rightarrow \max_{x \in D}; w \in W, \quad (6)$$

where  $W = \{w = (w_1, \dots, w_m)^T; w_i > 0; \sum_{i=1}^m w_i = 1\}$ .

The above allows one to formalize the decision-making process by reducing a set of criteria to a single target function, the value of which determines the optimal solution.

Factor weights  $w_i$  are identified with the numerical values of the corresponding utility functions. To select an alternative, the theorem of the method of multi-criteria utility theory is used, the essence of which is that if the criteria are independent in utility and preference, then there is a utility function.

$$U(x) = \sum_{i=1}^m w_i u_i(y_i), \quad (7)$$

which serves as a criterion for selecting the optimal option. At the same time  $U(x)$  is a multi-criteria utility function ( $0 \leq U(x) \leq 1$ ) for the alternative  $x$ ;  $u_i(y_i)$  is a utility function of the  $i$ -th criterion ( $0 \leq u_i(y_i) \leq 1$ );  $y_i$  is a value of the alternative  $x$  by the criterion  $i$ ;  $w_i$  is a weight of the  $i$ -th criterion, moreover  $0 < w_i < 1, \sum_{i=1}^m w_i = 1$ .

In general, the process of multi-criteria alternative selection in decision-making is based on the following basic assumptions [25]:

- the set of alternatives  $X$  is a finite set of elements  $X = \{x_1, x_2, \dots, x_n\}$  that the decision-maker can list;
- the evaluation of alternatives is carried out by utility functions  $f_i$ , moreover  $f_i: X \rightarrow R(i = 1, m)$ ;
- the decision-maker uses criteria ordered by their priority.

A set of criteria  $K_1, K_2, K_3, K_4$  is formed. The alternative options for the process are methods for evaluating the probability of information messages, denoted by M1, M2, M3. For each option,



$j$ -th alternative ( $j = 1, 2, 3$ ) by the  $i$ -th criterion ( $i = 1, \dots, 4$ ). Finally, the multi-criteria evaluation of the utility of the  $j$ -th alternative:

$$U_j = \sum_{i=1}^4 w_i u_{ij}; j = 1, 2, 3 \quad (8)$$

Additional notations are introduced:  $w_i$  is the initial weight of the  $i$ -th criterion, determined on the basis of an expert evaluation of its influence on the priority of the selected methods;  $g_{ij}$  are expertly established percentage values of the importance of the  $i$ -th criterion during the formation of the  $j$ -th alternative. It is important to note that the following condition must be met for each criterion:

$$\sum_{i=1}^4 g_{ij} = 100; j = 1, 2, 3 \quad (9)$$

The evaluation of alternatives by efficiency degrees  $g_{ij}$  is represented in Table 6.

**Table 6**

Evaluating alternatives of the Pareto set by criteria priority weights

Criteria	Weights of criteria	Evaluating the criteria efficiency in alternative options		
		M1	M2	M3
K1	$w_1$	$g_{11}$	$g_{12}$	$g_{13}$
K2	$w_2$	$g_{21}$	$g_{22}$	$g_{23}$
K3	$w_3$	$g_{31}$	$g_{32}$	$g_{33}$
K4	$w_4$	$g_{41}$	$g_{42}$	$g_{43}$

Considering the degrees of influence or efficiency of the criteria in different variants, that is, when applied to each of the selected methods, matrices of pairwise comparisons are constructed according to the method of hierarchy analysis. Processing these matrices allows one to obtain the corresponding utility functions  $u_{ij}$ , namely:  $K_1 - u_{11}, u_{12}, u_{13}$ ;  $K_2 - u_{21}, u_{22}, u_{23}$ ;  $K_3 - u_{31}, u_{32}, u_{33}$ ;  $K_4 - u_{41}, u_{42}, u_{43}$ .

An important addition to the previous considerations. It is obvious that the criteria of the Pareto set form a new autonomous group, which requires the calculation of current weight values  $w_i$  based on the initial numerical priorities. The obtained values will be used for the final calculation of the final target functions.

The formal representation of the pairwise comparison matrix of the initial weights is presented in Table 7.

**Table 7**

Pairwise comparison matrix of weights of the Pareto set criteria

Criteria	$K_1$	$K_2$	$K_3$	$K_4$
$K_1$	1	$w_1/w_2$	$w_1/w_3$	$w_1/w_4$
$K_2$	$w_2/w_1$	1	$w_2/w_3$	$w_2/w_4$
$K_3$	$w_3/w_1$	$w_3/w_2$	1	$w_3/w_4$
$K_4$	$w_4/w_1$	$w_4/w_2$	$w_4/w_3$	1

The sign "/" in the matrix elements means a comparison of the starting weight values of the criteria.

Processing the matrix will allow calculating the normalized weight values of the criteria, which will become the initial data for generating alternative options and determining the optimal method for evaluating the probability of text information messages.

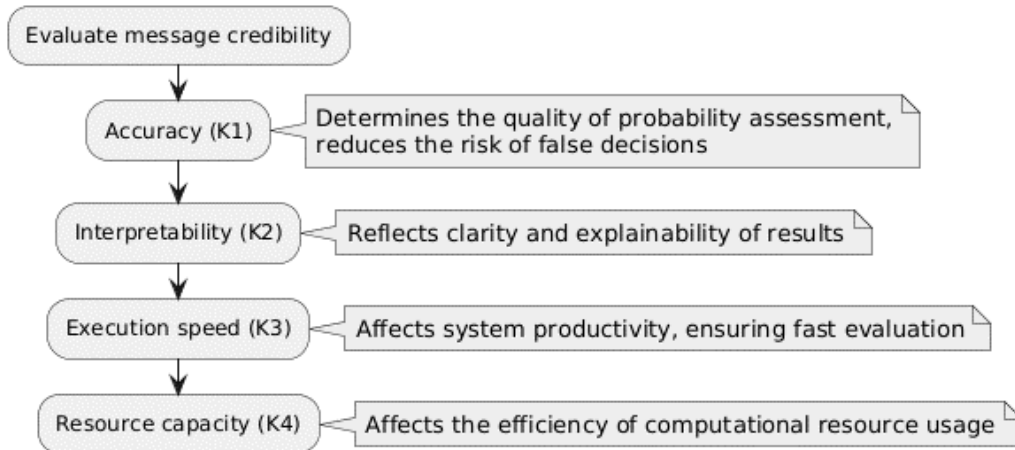
The final multi-criteria evaluations of the utility of alternatives for options M1, M2, M3, obtained on the basis of the formula (8), are expressed by the relation (10):

$$\begin{aligned} U_1 &= w_1 \cdot u_{11} + w_2 \cdot u_{21} + w_3 \cdot u_{31} + w_4 \cdot u_{41}; \\ U_2 &= w_1 \cdot u_{12} + w_2 \cdot u_{22} + w_3 \cdot u_{32} + w_4 \cdot u_{42}; \\ U_3 &= w_1 \cdot u_{13} + w_2 \cdot u_{23} + w_3 \cdot u_{33} + w_4 \cdot u_{43}. \end{aligned} \quad (10)$$

As noted earlier, the indicator of selecting the optimal method among alternative options for evaluating the probability of messages is the option (method) for which the value of the utility function of the combined partial target functions in the relation (6) reaches the maximum indicator.

Since the experimental implementation of the above theoretical approaches requires the information about the levels of the criteria preferences, a multi-level graphical model is designed (Fig. 3), which visually reproduces the essence of the criteria and the priority of the influence on the evaluation of text information messages. The preferences of the criteria in the model are reproduced on the basis of the diagram (Fig. 1).

To form the model, the essence of the criteria identified for the study are briefly summarized. Thus, methods for text message processing are characterized by: **accuracy**, which determines the ability of the method to correctly classify or analyse text data; **interpretability**, which reflects the clarity and explainability of the results, which is critically important in areas where transparency of decision-making is required; **execution speed**, which affects the performance of the system and depends on the algorithmic complexity of the calculations; **resource capacity**, which determines the amount of memory and computing power used, which is especially important when working with large amounts of text data.



**Figure 3:** Model of the priority influence of criteria on the message evaluation process.

Thus, an algorithm for generating and calculating alternative options is constructed, suitable for studying the process of message evaluation using modern methods and functional criteria. The essence of its theoretical foundations is based on mathematical calculations of modelling theory, methods of hierarchy analysis and decision-making, and operations research theory.

#### 4. Experiment, results and discussion

As part of further research, the refined normalized weighted priorities of the model criteria are calculated (Fig. 3), necessary for practical implementation (Table 6). To do this, using the scale of

relative importance of objects [27, 28], a pairwise comparison matrix of criteria is formed, taking into account the levels of their preference in the priority model.

**Table 8**

Pairwise comparison matrix of message evaluation criteria according to the model in Fig. 1

Criteria	$K_1$	$K_2$	$K_3$	$K_4$
$K_1$	1	2	3	4
$K_2$	1/2	1	2	3
$K_3$	1/3	1/2	1	2
$K_4$	1/4	1/3	1/2	1

Processing the matrix results in obtaining the initial numerical preferences of the criteria expressed in conventional units, namely:  $K_1$  (accuracy) – 220 c.u.;  $K_2$  (interpretability) – 130 c.u.;  $K_3$  (execution speed) – 80 c.u.;  $K_4$  (resource capacity) – 50 c.u. At the same time, normalized weight values of the criteria are calculated:  $w_1 = 0,46$ ;  $w_2 = 0,28$ ;  $w_3 = 0,16$ ;  $w_4 = 0,10$

As a result of processing the matrix, the following are obtained: the maximum eigenvalue of the matrix  $\lambda_{max}$ , the consistency index  $IU = 0,01$  and the consistency ratio  $WU = 0,01$ .

The calculation results correspond to the permissible limits of reliability, the essence of which is as follows. The assessment of the obtained solution is determined by the consistency index, the value of which is determined by the formula  $IU = \frac{(\lambda_{max}(n))}{(n-1)}$ , where  $n$  is a number of objects. The consistency index value is compared with the reference values of the consistency indicator, the so-called random index ( $WI$ ), which depends on the number of objects being compared. In this case, the results are considered satisfactory if the index value does not exceed 10% of the reference value  $WI$ . Comparing the obtained value  $IU$  and the reference value  $WI = 0,9$  for four criteria, and checking the inequality  $IU < 0,1 \times WI$ , one obtains:  $0,01 < 0,1 \times 0,9$ . This confirms the reliability of the obtained results. Additionally, the results are evaluated by the consistency ratio:  $WU = \frac{IU}{WI}$ . One obtains  $WU = 0,01$ . The results are considered satisfactory if  $WU \leq 0,1$ . Therefore, one has a sufficient level of process convergence and proper consistency of expert judgments regarding pairwise comparisons of criteria.

Normalized weight coefficients will be applied when calculating the utility functions of the criteria used in alternative approaches to message evaluation.

Additionally, possible groups of combinations of the efficiency shares of the criteria in alternative options are presented (Table 9), expressed in percentages [25,29].

**Table 9**

Combinations of efficiency shares of criteria in alternative options

Possible groups of combinations of the criteria efficiency			
10 - 10 - 80	20 - 10 - 70	30 - 10 - 60	40 - 10 - 50
10 - 20 - 70	20 - 20 - 60	30 - 20 - 50	40 - 20 - 40
10 - 30 - 60	20 - 30 - 50	30 - 30 - 40	40 - 30 - 30
10 - 40 - 50	20 - 40 - 40	30 - 40 - 30	40 - 40 - 20
10 - 50 - 40	20 - 50 - 30	30 - 50 - 20	40 - 50 - 10
10 - 60 - 30	20 - 60 - 20	30 - 60 - 10	50 - 10 - 40
10 - 70 - 20	20 - 70 - 10	60 - 10 - 30	50 - 20 - 30
10 - 80 - 10	70 - 10 - 20	60 - 20 - 20	50 - 30 - 20
80 - 10 - 10	70 - 20 - 10	60 - 30 - 10	50 - 40 - 10

Based on the data obtained, a basic table for the method of linear convolution of criteria is formed (Table 10).

**Table 10**

Experimental evaluation of alternatives by degrees of criteria efficiency

Criteria	Normalized criteria weights	Evaluating the criteria efficiency in alternative options in %		
		M1	M2	M3
Accuracy ( $K_1$ )	0,46	20	10	70
Interpretability ( $K_2$ )	0,28	50	30	20
Execution speed ( $K_3$ )	0,16	20	50	30
Resource capacity ( $K_4$ )	0,10	20	40	40

The following methods determine the alternative options in the table: M1 – logistic regression; M2 – decision trees; M3 – SVM (support vector machine).

One can calculate the values of the utility functions for each of the criteria. To do this, pairwise comparison matrices are formed that will reflect the preferences of alternatives relative to each criterion. The consistency of the obtained results is checked using the maximum eigenvalue of the priority vector for each matrix, the consistency index and the consistency ratio. To construct the matrices, the data from Table 10 are used, which contains the evaluations of the alternatives according to the selected criteria. The values of the alternatives are compared according to the criteria indicated in the last three columns of the table.

As a result, the pairwise comparison matrix for the criterion  $K_1$  "Accuracy" taking into account the shares of its efficiency in the alternative options will look like this (Table 11).

**Table 11**Pairwise comparison matrix for the criterion  $K_1$ 

$K_1$	M1	M2	M3
M1	1	2	1/4
M2	1/2	1	1/7
M3	4	7	1

The matrix analysis determines the following results:  $\lambda_{max}IU = 0,00$ ;  $WU = 0,00$ . Utility functions of the criterion "Accuracy"  $K_1$  are:  $u_{11} = 0,186$ ;  $u_{12} = 0,097$ ;  $u_{13} = 0,715$ .

The matrix for the criterion  $K_2$  "Interpretability" looks like this (Table 12).

**Table 12**Pairwise comparison matrix for the criterion  $K_2$ 

$K_2$	M1	M2	M3
M1	1	2	6
M2	1/2	1	5
M3	1/6	1/5	1

The resulting data are obtained:  $\lambda_{max} = 3,03$ ;  $IU = 0,01$ ;  $WU = 0,03$ . The utility functions of the criterion  $K_2$  "Interpretability" in alternative options are expressed by the following values:  $u_{21} = 0,576$ ;  $u_{22} = 0,341$ ;  $u_{23} = 0,081$ .

The criterion  $K_3$  "Execution speed" will form a similar matrix (Table 13).

**Table 13**Pairwise comparison matrix for the criterion  $K_3$ 

$K_3$	M1	M2	M3
M1	1	1/5	1/4
M2	5	1	3
M3	4	1/3	1

After the matrix analysis  $\lambda_{\max} = 3,08$ ;  $IU = 0,04$ ;  $WU = 0,07$ . The utility functions of the criterion "Execution speed" regarding alternatives are:  $u_{31} = 0,093$ ;  $u_{32} = 0,626$ ;  $u_{33} = 0,279$ .

The criterion  $K_4$  "Resource capacity" will result in the following pairwise comparison matrix (Table 14).

**Table 14**Pairwise comparison matrix for the criterion  $K_4$ 

$K_4$	M1	M2	M3
M1	1	3	3
M2	1/3	1	1
M3	1/3	1	1

As a result of processing the matrix, it is obtained:  $\lambda_{\max} = 3,00$ ;  $IU = 0,00$ ;  $WU = 0,00$ . The utility of the criterion "Resource capacity" is expressed by the following values:  $u_{41} = 0,6$ ;  $u_{42} = 0,2$ ;  $u_{43} = 0,2$ .

Calculation indicators, in particular the maximum values of priority vectors  $\lambda_{\max}$ , consistency indices  $IU$  and consistency ratio  $WU$ , meet the established requirements presented above. Substituting the weight coefficients and the values of the utility functions of the criteria into the relation (10), the final values of the combined functionalities of the alternative options are obtained (11).

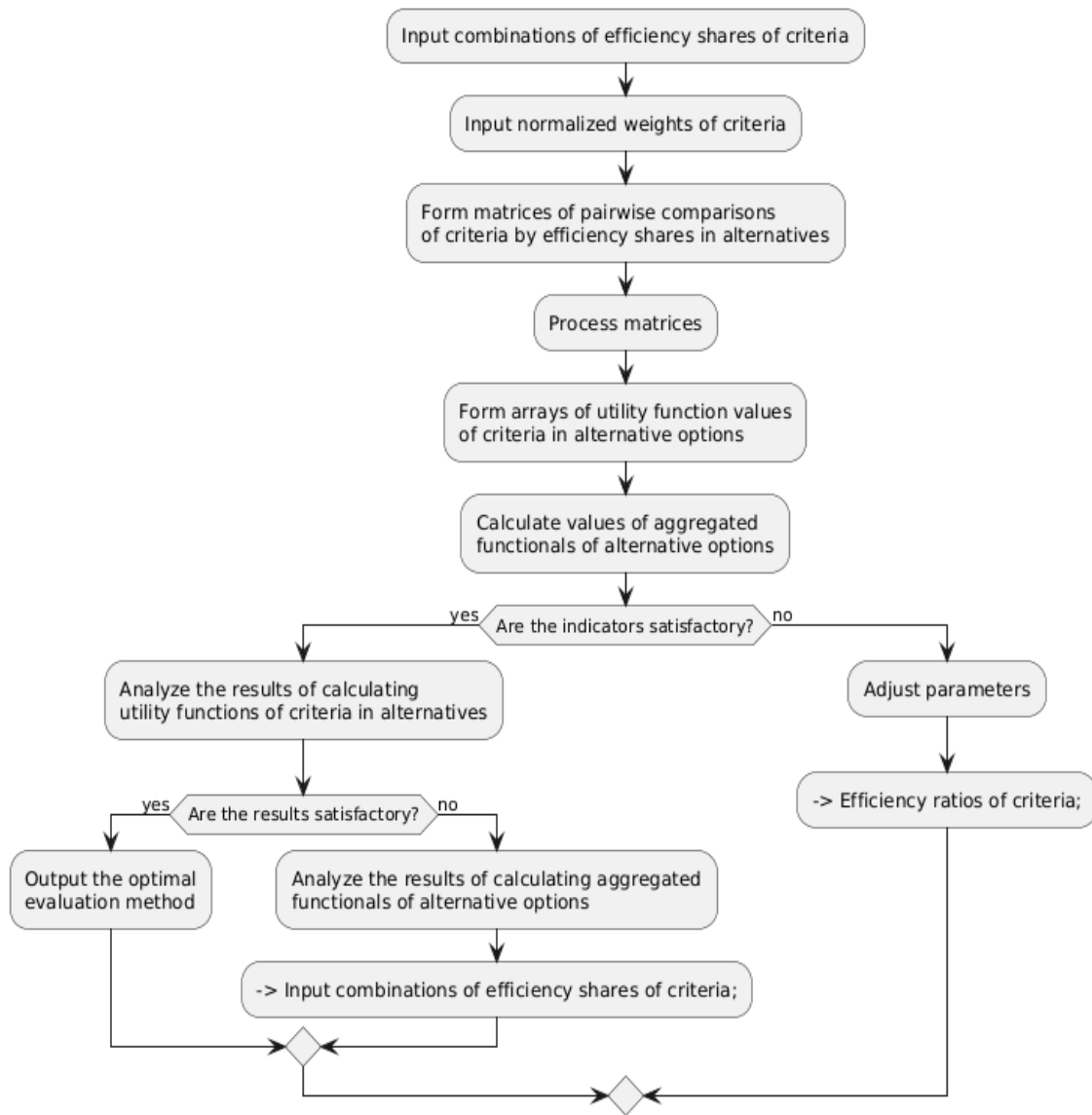
$$\begin{aligned}
 U_1 &= 0.46 \cdot 0.186 + 0.28 \cdot 0.576 + 0.16 \cdot 0.093 + 0.1 \cdot 0.6; \\
 U_2 &= 0.46 \cdot 0.097 + 0.28 \cdot 0.341 + 0.16 \cdot 0.626 + 0.1 \cdot 0.2; \\
 U_3 &= 0.46 \cdot 0.715 + 0.28 \cdot 0.081 + 0.16 \cdot 0.279 + 0.1 \cdot 0.2.
 \end{aligned} \tag{11}$$

Finally, one gets:  $U_1 = 0.322$ ;  $U_2 = 0.260$ ;  $U_3 = 0.416$ .

Among the selected alternative methods for evaluating the probability of information messages, the third option, i. e. the method M3 – SVM (support vector machine) is considered the most efficient, since it is characterized by the maximum value of the combined functional  $U_3$ . The theoretical justification of this option indicates its ability to provide a high level of reliability of the obtained score. Important parameters that determine the method efficiency are its ability to accurately evaluate the veracity of messages and efficiently distribute computational resources. It is the combination of such criteria as "accuracy" and "resource capacity" that makes this option the best among other alternatives.

The functional model of the algorithm of the probable software solution is shown in Fig. 4.

This model will take into account a multifactor analysis of the characteristics of input data and adaptively select the most effective approach depending on the context. It also ensures integration with machine learning systems to enhance the accuracy of assessment and reduce the rate of false-positive results. For this purpose, machine learning algorithms such as classification methods, regression, and neural networks are utilized, allowing the detection of hidden patterns in large data sets. Additionally, the model can apply reinforcement learning techniques to dynamically adjust evaluation criteria based on changing conditions in the information environment.



**Figure 4:** Functional model of the algorithm for automated determination of the optimal method for evaluating the probability of information messages.

Thus, based on the results of the conducted study, the most effective method for assessing the credibility of messages is the Support Vector Machine (SVM). This method is based on finding a hyperplane that maximally separates different data classes in the feature space, making it optimal for binary classification tasks, such as distinguishing reliable information from disinformation or manipulative content. However, the application of SVM for message credibility assessment also has certain limitations. One of the key challenges is the complexity of interpreting the results, as SVM does not provide clear explanations regarding which specific text features influenced the classification. Additionally, this method is sensitive to imbalanced datasets, where one category of messages significantly outweighs the other, which can distort classification results.

It should be emphasized that the final selection of the optimal option within the method of linear convolution of criteria can be improved by developing an appropriate software application, which will allow automating the process of calculating the values of combined functionals for various alternative methods and will provide the ability to take into account a wide range of options for combinations of criteria efficiency degrees.

The application of machine learning methods in this process is extremely important, as they enable the analysis of large volumes of data, the identification of hidden dependencies between criteria, and the automatic adjustment of weight coefficients to improve assessment accuracy.

Furthermore, the use of machine learning will facilitate the model's adaptation to changing conditions and enhance its ability to self-learn, ensuring a more flexible and reliable approach to selecting the optimal evaluation method. A promising direction for future research is the refinement of the proposed approach through the integration of advanced artificial intelligence methods and hybrid algorithms that combine classical statistical models with neural network technologies. This will improve the system's adaptability to dynamic changes in the information environment, increase its resistance to manipulative influences, and provide a more accurate and well-founded assessment of message credibility in real-time.

In conclusion, it is worth emphasizing that the research is based on the concept of the relationship between the selection of the optimal method for evaluating the probability of information messages and the combination of efficiency shares of the applied criteria. This approach makes it possible to identify patterns that affect the accuracy and validity of the intellectual analysis of manipulative content in the information space. The recommended algorithmic decision to the automated solution of the problem provides an increase in the level of the evaluation objectivity, reducing the influence of subjective factors. In addition, the proposed solution provides the possibility of flexible parameter settings, which allows adapting the system to the specific conditions and requirements of a specific research task, contributing to increasing its efficiency and practical value.

## 5. Conclusions

The results of the study confirm the relevance of the problem of evaluating methods of intellectual analysis and detecting manipulative content in the information space. Determining the most efficient method is a key task for ensuring the information security and countering disinformation. Given the constantly growing threats in the information environment, combating manipulation and disinformation requires not only the improvement of technical analysis methods but also the development of strategies. This includes educational programs, increasing media literacy among the population, and active collaboration between government institutions and scientific organizations, which will enable effective resistance to manipulative content and ensure sustainable information security.

The study proposes an approach based on the formation of a Pareto set of mutually non-dominated criteria, which allows one to objectively evaluate the efficiency of machine learning methods, in particular logistic regression, decision trees and the support vector method (SVM). The generation of alternatives is carried out by identifying options determined by the selected methods taking into account expertly established efficiency degrees of key criteria, such as accuracy, execution speed, resource capacity and interpretability.

The application of the method of hierarchies analysis to the developed model of the priority influence of criteria on the process of analysis, evaluation and detection of manipulative content provides the calculation of the utility functions of the criteria for each of the methods. As a result, combined functionalities of alternative options are formed, which act as the main optimization criterion. The use of multi-criteria analysis methods, in particular the method of linear convolution of criteria, provides a balanced approach to selecting the optimal method for evaluating the probability of information messages, contributing to increasing the accuracy and validity of decisions made.

The results of the study indicate that the selection of the optimal method largely depends on the specific requirements for the evaluation system. In particular, in cases where accuracy is the priority criterion, it is advisable to use the support vector method (SVM), while to ensure high data processing speed it is more expedient to use decision trees. Logistic regression, in turn, provides a balance between accuracy and interpretability, which makes it an efficient tool for analysing text messages.

In view of the prospects for further development, the approach proposed in the study makes it possible to adapt the evaluation system to the changing conditions of the information environment, integrate additional analysis criteria and use modern machine learning algorithms to automate the content evaluation process. The results obtained can be used to develop automated information space

monitoring systems, which will contribute to the timely detection of manipulative content and increase the level of information security.

The proposed approaches can be integrated into software solutions for intelligent analysis and evaluation of the level of manipulative content in the information space, which is especially relevant in the context of the growing information influence on the public opinion and political processes. Further research should be directed at expanding the approaches by using the latest methods, in particular deep learning and hybrid models, which can potentially provide significantly higher efficiency in detecting manipulative content.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] E. Aïmeur, S. Amri, G. Brassard. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1) (2023) 30. doi:10.1007/s13278-023-01028-5.
- [2] L. Chen, J. Chen, C. Xia. Social network behavior and public opinion manipulation. *Journal of Information Security and Applications*, 64 (2022) 103060. doi:10.1016/j.jisa.2021.103060.
- [3] M. Basol, J. Roozenbeek, S. Linden. Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, 3(1) (2020) 2. doi:10.5334/joc.91.
- [4] G. Rampersad, A. Turki. Fake news: Acceptance by demographics and culture on social media. *Journal of Information Technology & Politics*, 17(1) (2020) 1–11. doi:10.1080/19331681.2019.1686676.
- [5] S. Jones-Jang, T. Mortensen, J. Liu. Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2) (2021) 371–388. doi:10.1177/0002764219869406.
- [6] M. Ozkan-Okay, E. Akin, Ö. Aslan, S. Kosunalp, T. Iliev, I. Stoyanov. A comprehensive survey: Evaluating the efficiency of artificial intelligence and machine learning techniques on cyber security solutions. *IEEe Access*, 12 (2024) 12229–12256. doi:10.1109/ACCESS.2024.3355547.
- [7] D. Caled, M. J. Silva. Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science*, 5 (2022) 123–159. doi:10.1007/s42001-021-00118-8.
- [8] S. R. Sahoo, B. B. Gupta. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, 100 (2021) 106983. doi:10.1016/j.asoc.2020.106983.
- [9] S. Mishra, P. Shukla, R. Agarwal. Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, 2022(1) (2022) 1575365. doi:10.1155/2022/1575365.
- [10] P. Dhiman, A. Kaur, C. Iwendi, S. K. Mohan. A scientometric analysis of deep learning approaches for detecting fake news. *Electronics*, 12(4) (2023) 948. doi:10.3390/electronics12040948.
- [11] T. Hovorushchenko, V. Alekseiko, V. Shvaiko, J. Ilchyshyna, A. Kuzmin. Information system for earth's surface temperature forecasting using machine learning technologies. *Computer Systems and Information Technologies*, 4 (2024) 51–58. doi:10.31891/csit-2024-4-7.
- [12] Z. Khanam, B. N. Alwasel, H. Sirafi, M. Rashid. Fake news detection using machine learning approaches. *IOP Conference Series: Materials Science and Engineering*, 1099(1) (2021) 012040. doi:10.1088/1757-899X/1099/1/012040.
- [13] A. Naitali, M. Ridouani, F. Salahdine, N. Kaabouch. Deepfake attacks: Generation, detection, datasets, challenges, and research directions. *Computers*, 12(10) (2023) 216. doi:10.3390/computers12100216.



- [14] D. Michail, N. Kanakaris I. Varlamis. Detection of fake news campaigns using graph convolutional networks. *International Journal of Information Management Data Insights*, 2(2) (2022) 100104. doi:10.1016/j.jjimei.2022.100104.
- [15] J. Jing, H. Wu, J. Sun, X. Fang, H. Zhang. Multimodal fake news detection via progressive fusion networks. *Information Processing & Management*, 60(1) (2023) 103120. doi:10.1016/j.ipm.2022.103120.
- [16] N. Capuano, G. Fenza, L. Vincenzo, F. Nota. Content-based fake news detection with machine and deep learning: a systematic review. *Neurocomputing*, 530 (2023) 91–103. doi:10.1016/j.neucom.2023.02.005
- [17] M. A. K. Raiaan, S. Sakib, N. M. Fahad, A. Rahman, S. Shatabda, H. Mukta. A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision Analytics Journal*, (2024) 100470. doi:10.1016/j.dajour.2024.100470.
- [18] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, H. Malik. Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, 53(4) (2023) 3974–4026. doi:10.1007/s10489-022-03766-z.
- [19] T. R. McIntosh, T. Susnjak, T. Liu, P. Watters, M. N. Halgamuge. The inadequacy of reinforcement learning from human feedback-radicalizing large language models via semantic vulnerabilities. *IEEE Transactions on Cognitive and Developmental Systems*, (2024) 1561–1574. doi:10.1109/TCDS.2024.3377445.
- [20] M. Carroll, A. Chan, H. Ashton, D. Krueger. Characterizing manipulation from AI systems. 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, (2023) 1–13. doi:10.1145/3617694.362322.
- [21] S. Mahajan, A. Chauhan, S. K. Gupta. On Pareto optimality using novel goal programming approach for fully intuitionistic fuzzy multiobjective quadratic problems. *Expert Systems with Applications*, 243 (2024) 122816. doi:10.1016/j.eswa.2023.122816.
- [22] N. Joshi, R. Sudeep, R. Bapat, R. N. Sengupta. Optimal estimation of reliability parameter for inverse Pareto distribution with application to insurance data. *International Journal of Quality & Reliability Management*, (2024) 811–1837. doi:10.1108/IJQRM-06-2023-0213.
- [23] J. Christensen, S. Hesstvedt. The influence of expert groups: a citation analysis. *Journal of European Public Policy*, 31(5) (2024) 1259–1294. doi:10.1080/13501763.2023.2174168.
- [24] Z. Chonghui, X. Dong, S. Zeng, L. A. Carlos. Dual consistency-driven group decision making method based on fuzzy preference relation. *Expert Systems with Applications*, 238 (2024) 122228. doi:10.1016/j.eswa.2023.122228.
- [25] V. Senkivskyy, I. Pikh, S. Havenko, S. Babichev. A model of logical inference and membership functions of factors for the printing process quality formation. *Lecture Notes in Computational Intelligence and Decision Making: Proceedings of the XV International Scientific Conference ISDMCI'2019, Ukraine, May 21–25, 15 (2019) 609–621*. doi:10.1007/978-3-030-26474-142.
- [26] V. Senkivskyy, A. Kudriashova, I. Pikh, I. Hileta, O. Lytovchenko. Models of postpress processes designing. 1st International Workshop on Digital Content & Smart Multimedia, DCSMart 2019, Lviv, Ukraine, December 23–25, (2019) 259–270. URL: <https://ceur-ws.org/Vol-2533/paper24.pdf>.
- [27] C. Dumitrescu, P. Ciotirnae, C. Vizitiu. Fuzzy logic for intelligent control system using soft computing applications. *Sensors*, 21(8) (2021) 2617. doi:10.3390/s21082617.
- [28] S. Babichev, B. Durnyak, V. Zhydetsky, I. Pikh, V. Senkivskyy. Techniques of DNA microarray data pre-processing based on the complex use of bioconductor tools and shannon entropy. *CEUR Workshop Proceedings 2353, Zaporizhzhia, April 19, (2019) 365–377* URL: <https://ceur-ws.org/Vol-2353/paper29.pdf>.
- [29] I. Ogirko, M. Dubnevych, T. Holubnyk, N. Pysanchyn, O. Ogirko. Information technology to determine the efficiency of advertising communication. *IEEE International Scientific-Practical Conference “Problems of Infocommunications. Science and Technology” (PIC S&T-2018), Kharkiv, Ukraine, (2018) 752–756*. doi:10.1109/INFOCOMMST.2018.8632016.