UDC 004.5:355

D. Dzhendzhero **V. Nakonechnyi** **A. Poberezhnyi**

METHODOLOGICAL APPROACHES TO IDENTIFYING THE FEATURES OF FALSE INFORMATION WITHIN THE INFORMATION SECURITY SYSTEM OF THE DEFENSE FORCES OF UKRAINE

The methodological foundations for identifying indicators of false information within the information security system of the Defense Forces of Ukraine are substantiated. Contemporary scientific approaches to disinformation identification are analyzed, including cognitive, content-semantic, behavioral, technical, and source-based approaches. It is established that existing studies tend to focus on individual aspects of the problem, while the integration of these approaches into a unified methodological framework remains insufficiently developed. A classification of five groups of false-information indicators and methods for their quantitative assessment are proposed. An integral model for assessing information credibility is developed, enabling a transition from qualitative description to a formalized quantitative evaluation.

Keywords: *hybrid warfare, disinformation, fake news, information credibility, information security, features of false information, credibility assessment model.*

Statement of the problem. Hybrid warfare represents one of the most complex and multidimensional forms of modern confrontation, combining military, political, economic, information-psychological, and cybernetic means of influence aimed at destabilizing public administration systems and undermining national resilience.

In the context of ongoing hybrid aggression against Ukraine, information security within the Defense Forces acquires strategic importance. The adversary actively employs disinformation and psychological manipulation technologies to distort the operational environment, construct false narratives regarding combat actions, losses, logistics, and troop morale. Such information attacks are designed to destabilize command systems, reduce trust in leadership, and create decision-making chaos. Therefore, identifying objective features of false information has direct practical significance for the information security system of the Defense Forces, as it enhances the reliability of operational information analysis and ensures the informational resilience of military units.

The information component of hybrid warfare has acquired a systemic nature, encompassing the deliberate use of disinformation, propaganda, and simulated fake messages that influence mass consciousness, distort reality, and create a favorable background for political and military decisions of the adversary [1].

Studies of modern conflicts show that hybrid warfare is based on a "synergy of threats" – the simultaneous application of military, informational, and sociotechnical methods of influence, where the informational component determines the speed and effectiveness of operations [2]. This combination makes such wars particularly dangerous for democratic states, as the boundary between traditional aggression and information-cognitive influence becomes nearly invisible. Consequently, classical cybersecurity measures focused solely on technical parameters prove insufficient.

At the same time, the global information space demonstrates exponential growth in the scale and speed of false information dissemination. Analyses of modern disinformation campaigns reveal that they develop according to epidemic-like patterns – fake news spreads faster than fact-checking processes occur, and repeated replication creates an illusion of credibility [3]. As a result, a new class of semantic risks emerges that cannot be detected solely by technical monitoring means.

Thus, the relevance of this study is determined by the growing role of disinformation as a key instrument of hybrid aggression, the increasing synergy of modern threats, and the need to develop comprehensive methodological approaches for identifying the features of false information that would enable systematic detection and counteraction. The development of such approaches is a prerequisite for enhancing the efficiency of information resource protection and strengthening the state's informational resilience.

Analysis of recent research and publications. The issue of defining the features of false information in the context of hybrid warfare is increasingly regarded not only as a communication or social problem but also as a methodological task of identifying and classifying informational influences. Ukrainian scholars propose various approaches to analyzing the phenomenon of disinformation; however, most works remain within specific disciplinary boundaries – law, social sciences, or communications – without forming an integrated system of criteria.

Study [4] examines a cognitive-communication approach to detecting false messages, emphasizing behavioral trust indicators and patterns of content dissemination in social media. It highlights that users' cognitive reactions can serve as indicators of fake materials, as emotional intensity and word repetition create an illusion of credibility.

From a legal perspective, study [5] examines a normative-evaluative approach to defining false information. The authors analyze existing legislative mechanisms and practices for countering disinformation, demonstrating that current regulation is primarily focused on the application of sanctions and lacks clearly formulated methodological criteria for assessing information credibility. At the same time, the need to form unified signs of falsehood, which could be integrated into the state's information security system, is emphasized.

In the international context, the NATO analytical report [6] presents a multidisciplinary model for evaluating information credibility. It involves the cooperation of military, scientific, and civil institutions, integrating technical, semantic, and cognitive methods of message analysis. The report defines fundamental principles of data verification such as contextual assessment, source validation, and identification of recurring informational patterns.

The methodological evolution is clearly traced in study [7], where disinformation is examined as part of the cognitive-narrative approach to hybrid warfare. The authors introduce the concept of strategic narrative warfare, according to which the features of false information are identified through semantic structure analysis, inconsistencies between facts, and narrative coherence. This approach enables the detection of informational attacks at the level of meaning rather than mere factual accuracy.

In study [8], content-semantic and behavioral approaches to detecting false information on social networks are systematized. The authors summarize linguistic, semantic, and source-based indicators, as well as content dissemination patterns that may indicate the manipulative or disinformational nature of messages. The study emphasizes the advisability of combining these indicators to improve the accuracy of automated detection methods.

The cognitive-behavioral approach to disinformation detection is presented in study [9], focusing on processes of perception and trust. The authors argue that the features of false information manifest not only in content but also in its mode of consumption – particularly in dissemination speed, uniformity of user reactions, and shifts in the emotional tone of communication. This approach serves as the foundation for developing cognitive models of disinformation risk.

In recent works by security and defense experts, increasing attention is devoted to the issue of information credibility within the process of operational and combat management. It is emphasized that the informational component of defense operations must rely on verified data and real-time validation mechanisms. Considering the experience of defense structures and the growing role of information analytics, this underscores the need to develop analytical models capable of distinguishing disinformation from operationally reliable data – a factor directly influencing the effectiveness of planning, coordination, and decision-making in the military command system.

In summary, modern research tends to focus on isolated approaches – legal, cognitive, content-semantic, or technical. At the same time, their integration into a unified methodological system for identifying features of false information remains insufficiently realized. This gap defines the scientific niche within which the present study is conducted.

The purpose of the article is to develop the conceptual foundations for forming methodological approaches to identifying the features of false information under the conditions of hybrid warfare.

The study aims to systematize existing scientific approaches, identify their strengths and weaknesses, and substantiate the need for integrating cognitive, semantic, legal, and technical criteria into a unified methodological framework.

To achieve this goal, the following main tasks are defined:

– to analyze modern approaches to the classification and detection of false information applied in the field of information security;

– to identify the key features of false messages that are relevant under conditions of hybrid informational influence;

– to substantiate the conceptual principles for constructing a comprehensive methodology for identifying false information, aimed at improving the efficiency of information resource protection systems.

Summary of the main material. Within the context of modern hybrid warfare, identifying the features of false information requires not only a descriptive but also a formalized approach that allows message verification with a defined level of accuracy. Based on the conducted analysis, a conceptual model has been developed that combines the classification of features, methods of their quantitative assessment, and an integral indicator of information credibility. This structure enables the transition from qualitative characteristics to measurable parameters, thereby increasing the objectivity of false message detection.

Classification of False Information Features. The systematization of features serves as the fundamental stage in building a methodology for determining falsehood. During the study, five principal groups of features were identified, covering the semantic, linguistic, behavioral, technical, and source-based aspects of information:

1. Content (semantic) features – characterize the meaning structure of the message, the logical coherence of statements, the level of emotional saturation, and the use of manipulative techniques. These include polarized vocabulary, appeals to fear or outrage, exaggerations, and excessive generalizations.

2. Linguistic features – are determined by the statistical characteristics of the text, such as the frequency of repeated words, the average sentence length, the level of entropy, and compliance with Zipf 's law.

For natural texts, the following relationship holds:

$$f(r) = \frac{C}{r^a}, \qquad (1)$$

where $f(r)$ is the frequency of a word with rank r;

C is the normalization coefficient;

a ≈ 1.

A significant deviation from this relationship indicates artificiality or manipulation within the text.

3. Behavioral features – reflect the pattern of message dissemination within the network. False information is typically characterized by a sharp increase in the number of reposts over a short period, synchronized publication of identical texts on different platforms, and a lack of thematic diversity in comments.

4. Technical features – include parameters of the message structure and the properties of the distribution source (metadata, timestamps, IP addresses, and interconnections between accounts). These features form the basis for constructing network graphs of information flows.

5. Source-based features – characterize the degree of credibility of the information source, its history of activity, and the ratio between factual and emotional content.

Methods for Determining the Features of False Information. For each group of features, an appropriate quantitative evaluation method is applied, allowing qualitative observations to be transformed into measurable parameters.

Entropy Analysis. To assess the structural complexity of text, Shannon's entropy is used, which determines the degree of linguistic disorder:

$$H = -\sum_{i=1}^{n} p_i log_2 p_i\,, \qquad (2)$$

where p_i is the probability of occurrence of the i-th word or symbol.

Natural texts exhibit an optimal level of entropy; excessively high or low entropy indicates artificial text construction (automatic generation, deliberate simplification, or templating).

Frequency Analysis and Zipf 's Law. If a text does not follow Zipf 's law, it indicates artificiality or lexical imbalance. Comparing the empirical frequency distribution with the theoretical one helps identify informational anomalies typical of fake messages.

Term Weighting (TF-IDF). To quantify the semantic significance of terms, the following formula is used:

$$w_{i,j} = tf_{i,j} \times log\,\frac{N}{df_i}\,, \qquad (3)$$

where $tf_{i,j}$ is the frequency of term i in document j;

df_i is the number of documents containing term i;

N is the total number of documents.

The resulting weights are used to detect thematic deviations and assess narrative similarity between sources.

Content Similarity Evaluation. To analyze message similarity within a single information campaign, the cosine similarity metric is applied:

$$cos\ (\theta) = \frac{A \cdot B}{||A||\ ||B||}, \qquad (4)$$

where A and B are feature vectors of two texts.

A high $cos\ (\theta)$ value (close to 1) indicates a significant similarity between messages, which implies coordinated dissemination of content from a common source.

Behavioral Modeling. Behavioral indicators are derived through the analysis of user activity time series. The replication rate coefficient is used to estimate the speed of message dissemination:

$$K_r = \frac{N_t - N_{t-1}}{\Delta t}, \qquad (5)$$

where N_t is the thenumber of publications or reactions at time t;

Δt is the observation interval.

A rapid increase in K_r over a short period indicates artificially stimulated content propagation.

Integral Model of Information Credibility Assessment. To generalize the analysis results, an integral model for determining message credibility is proposed, based on weighted summation of partial coefficients for each group of features:

$$D = \sum_{i=1}^{n} w_i k_i\ , \qquad (6)$$

where D is the integral credibility indicator;

w_i is the weight coefficient of the i-th feature's significance;

k_i is the normalized credibility value for that feature.

The w_i values are determined either by expert judgment or from statistical samples.

Each of the five groups of features is evaluated on a scale [0;1], where *1* corresponds to a fully credible message and *0* to typical disinformation. This approach enables quantitative comparison of messages by degree of truthfulness and detection of potentially harmful information injections.

The results are interpreted using the following scale:

– $D > 0.75$ – information is credible;

– $0.45 < D \leq 0.75$ – information is potentially manipulative and requires verification;

– $D \leq 0.45$ – information is false or significantly distorted.

The resulting indicator can be used for automated monitoring of the information space, as well as for decision-making in cybersecurity systems and analytical platforms for information security.

Thus, the proposed methodology enables comprehensive detection of false information under hybrid warfare conditions by combining content, linguistic, behavioral, and technical analyses. Its application enhances the efficiency of information resource protection, strengthens the informational resilience of the state, and creates a foundation for further automation of disinformation identification processes.

Conclusions

As a result of the conducted research, the conceptual foundations of a methodology for identifying the features of false information under hybrid warfare conditions have been developed. Based on the analysis of scientific sources and existing practices of disinformation identification, it has been established that most current approaches focus on separate aspects – cognitive, semantic, legal, or technical – and therefore fail to ensure a comprehensive assessment of message credibility. This confirmed the need to develop an integrated system of criteria that would combine multiple levels of analysis.

The proposed methodology is grounded in a systemic approach and involves the identification of five groups of features: cognitive-semantic, linguistic, behavioral, technical, and source-based. For each group, a set of parameters has been defined that allows for the quantitative assessment of credibility levels. An integral model has also been developed to calculate the overall truthfulness coefficient of a message. This enables the transition from descriptive analysis to a formalized evaluation of informational influences.

The practical significance of the obtained results lies in establishing a scientific and methodological foundation for the further design of algorithms and automated systems for disinformation detection. The proposed model can be utilized to enhance the level of information security of the Defense Forces of Ukraine. Integrating the developed model into information support systems and analytical divisions will allow early detection of destructive informational influences, reduce the risk of decision-making based on distorted data, and strengthen the informational resilience of units during operational and combat activities.

Future research should focus on refining the weighting coefficients of feature groups using machine learning methods, extending the model to multimedia content, and studying the dynamics of audience cognitive reactions as one of the key factors in the spread of disinformation.

References

1. Varga T. M. (2024). *Dezinformatsiia ta propahanda yak instrumenty vedennia hibrydnoi viiny rosii proty Ukrainy* [Disinformation and Propaganda as Tools of Hybrid Warfare Conducted by russia Against Ukraine]. Kyiv : NUOU [in Ukrainian].

2. Vuković J., Matika D., & Barić S. (2022). Hybrid Warfare Challenges. Zagreb : Croatian Defence Academy [in English].

3. Patel S. S., Maloney E. J., & Omer S. B. (2020). The Landscape of Disinformation on Health Crisis Communication During the COVID-19 Pandemic in Ukraine. *Journal of Global Health*, vol. 10 (2), pp. 1–8. DOI: https://doi.org/10.7189/jogh.10.020310 [in English].

4. Ratushna I. (2024). *Kohnityvno-komunikatsiini mekhanizmy spryiniattia dezinformatsii v tsyfrovomu seredovyshchi* [Cognitive and Communicative Mechanisms of Disinformation Perception in the Digital Environment]. *Informatsiine suspilstvo*, no. 2 (16), pp. 45–53 [in Ukrainian].

5. Smotrych D., Ivanov N. (2023). *Pravovi aspekty borotby z dezinformatsiieiu v Yevropeiskomu Soiuzi: uroky dlia Ukrainy* [Legal aspects of combating disinformation in the European Union: lessons for Ukraine]. *Visnyk Natsionalnoho universytetu "Lvivska politekhnika". Seriia: yurydychni nauky*, vol. 4 (40), pp. 155–161. DOI: https://doi.org/10. 23939/ law2023.40.155 [in Ukrainian].

6. NATO StratCom COE. (2024). Disinformation Resilience Index: A Multidisciplinary Model of Credibility Assessment. Riga : NATO Strategic Communications Centre of Excellence [in English].

7. Bachmann S. D., Gunneriusson H., & Giegerich B. (2023). Hybrid Warfare and Disinformation: A Ukraine War Perspective. *Defence Strategic Studies Journal*, vol. 12 (1), pp. 34–51. DOI: https://doi.org/10.2478/dssj-2023-0003 [in English].

8. Shen H., Li T., & Zhang, Y. (2023). Fake News Detection on Social Networks: A Survey. *Applied Sciences (MDPI)*, vol. 13 (21). DOI: https://doi.org/10.3390/app132111877 [in English].

9. Danyk Y., & Briggs C. (2023). Modern Cognitive Operations and Hybrid Warfare. *Security and Defence Quarterly*, vol. 42 (2), pp. 55–68. DOI: https://doi.org/10.35467/ sdq/168741 [in English].

The article was submitted to the editorial office 17.11.2025

УДК 004.5:355

Д. С. Дженджеро, В. С. Наконечний, А. А. Побережний

МЕТОДОЛОГІЧНІ ПІДХОДИ ДО ВИЗНАЧЕННЯ ОЗНАК НЕПРАВДИВОЇ ІНФОРМАЦІЇ В СИСТЕМІ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ СИЛ ОБОРОНИ УКРАЇНИ

Обґрунтовано концептуальні і методологічні засади визначення ознак неправдивої інформації в системі інформаційної безпеки сил оборони України. Акцентовано на зростанні значення інформаційної безпеки і когнітивної стійкості в сучасних конфліктах. Автори систематизують і критично аналізують застосовувані наразі підходи до виявлення дезінформації, інтегруючи когнітивний, контентно-семантичний, поведінковий, технічний та джерельний виміри. Установлено, що більшість проведених досліджень охоплюють лише окремі аспекти цього явища, не маючи єдиної методологічної рамки, здатної поєднати якісні і кількісні параметри оцінювання достовірності інформації.

Запропонований підхід передбачає комплексну класифікацію п’яти груп ознак неправдивої *інформації та визначає конкретні кількісні метрики для їхнього оцінювання. До них належать: ентропійний і частотний аналіз текстових даних, закон Ципфа для виявлення закономірностей лексичного розподілу, TF-IDF-оцінка семантичної значущості, метрика косинусної подібності для міжтекстового порівняння та коефіцієнт темпу реплікації для моделювання динаміки поширення повідомлень. Інтеграція цих компонентів формує інтегральну модель достовірності, яка забезпечує перехід від описових оцінок до формалізованої, вимірюваної перевірки правдивості інформації.*

Розроблена методологічна основа дає змогу здійснювати багаторівневий аналіз інформаційного контенту та закономірностей його поширення, забезпечуючи більш об'єктивну базу для виявлення маніпулятивних або дезінформаційних наративів у межах гібридних інформаційних впливів. Отримані результати сприяють розвитку наукового розуміння механізмів дезінформації та створюють підґрунтя для розроблення автоматизованих аналітичних систем, здатних виявляти гібридні інформаційні загрози в режимі реального часу.

Практична значущість дослідження полягає у можливості застосування його результатів для проєктування інтелектуальних систем моніторингу інформаційного простору та механізмів раннього попередження про дезінформаційні кампанії.

Перспективними напрямами подальших досліджень є уточнення вагових коефіцієнтів індикаторів достовірності з використанням методів машинного навчання, поширення моделі на мультимедійний контент, а також дослідження динаміки когнітивних реакцій аудиторії як ключового чинника поширення неправдивої інформації.

Ключові слова: *гібридна війна, дезінформація, фейкові повідомлення, достовірність інформації, інформаційна безпека, ознаки неправдивої інформації, модель оцінювання достовірності.*

Dzhendzhero Dmytro – Postgraduate Student of the Department of Cybersecurity and Information Protection, Taras Shevchenko National University of Kyiv

https://orcid.org/0009-0007-9999-850X

Nakonechnyi Volodymyr – Doctor of Technical Sciences, Professor, Professor of the Department of Cybersecurity and Information Protection, Taras Shevchenko National University of Kyiv

https://orcid.org/0000-0002-0247-5400

Poberezhnyi Andrii – Researcher of the Scientific Research Laboratory of Service and Combat Application of the NGU, National Academy of the National Guard of Ukraine

https://orcid.org/0000-0002-8984-6912