Security
of Infocommunication Systems and Internet of Things

2024 Vol 2, No 2

https://doi.org/10.31861/sisiot2024.2.02004

Received 11 November 2024; revised 05 December 2024; accepted 20 December 2024; published 30 December 2024

Risk Management of Information Threats in IT with the Help of Intelligent Disinformation Detection Systems

Dmytro Uhryn[1,*], Yuriy Ushenko[1], Myroslav Kovalchuk[1] and Mykyta Zakharov[1]

[1]Computer Science Department, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

*Corresponding author (E-mail: d.ugryn@chnu.edu.ua)

ABSTRACT The article describes an intelligent disinformation detection system that counteracts the spread of false information on the Internet. It uses modern text and data processing methods, including machine learning and natural language processing, to accurately identify fake news. The main function of the system is to provide a predictive assessment of the reliability of information, which helps users make informed decisions, minimising the risk of falling under the influence of disinformation. Real-world testing of the system has confirmed its ability to quickly identify fake information, contributing to the information literacy of users and raising their awareness of information threats. Such a system is an important element in strengthening information security, which is especially important for Ukraine in the face of numerous information challenges. The system also plays an important role in managing the risks of information threats in IT. The results of the study made it possible to identify potential threats in the form of disinformation that could be used to manipulate public opinion or undermine trust in institutions. Integration of intelligent systems into risk management processes allows for a timely response to threats, reducing their impact on IT infrastructure and preserving the reputation of organisations. The system can be applied not only in the IT sector, but also in journalism, education, and public administration, where it helps prevent disinformation that has a serious social impact. It can also be used to monitor and analyse information flows, which helps identify and counteract false information. Thus, the developed system is an important step in strengthening information security, providing protection against fake news and serving as an effective tool for managing the risks of information threats in today's digital society.

KEYWORDS intelligent system, disinformation detection, artificial intelligence, machine learning, deep learning.

I. INTRODUCTION

In today's interconnected world, disinformation has become a serious global problem that has a significant impact on society, the economy and politics. With the proliferation of the Internet and social media, false information is transmitted at a high speed, which underscores the importance of detecting and regulating it. Disinformation can take many forms, from the deliberate spread of false news to the distortion of facts that can influence public opinion, political processes and the stability of society.

In today's context, researching and countering disinformation is becoming increasingly important. This study aims to reduce the negative effects of disinformation, promote informed decision-making and raise public awareness. Timely and accurate detection of false information allows for more effective management of social, cultural and economic processes that disinformation can disrupt.

The main goal of the study is to create an intelligent system for detecting disinformation that uses the latest machine learning and text analysis technologies. Such a system should be able to process large amounts of data, identify potentially false messages, and provide users with tools to quickly respond to disinformation.

The research focuses on disinformation in the online space, including social media, news sites and other online resources. The project examines disinformation distribution models, factors that contribute to its spread, and ways to identify it. The research focuses on the development of an intelligent system for detecting disinformation. The project combines machine learning, text categorisation and natural language processing techniques to identify false information in real time.

The project is of practical importance in the areas of information policy, strategic planning and security. The system helps raise public awareness and provides tools for reasoned discussion of disinformation issues. The project's intelligent algorithms provide useful information for decision-making and a deeper understanding of the spread of false information.

II. LITERATURE REVIEW

The development of a system for detecting disinformation is extremely important and justified in the current context. Traditional text analysis methods often have significant limitations [1], such as high processing time, limited access to data, and lack of real-time information. These shortcomings make it difficult to gain a deep understanding of information flows and make informed decisions.

Disinformation is a complex phenomenon that is influenced by numerous factors, including social, political and economic aspects. It is important for governments, organisations and policymakers to analyse these trends in order to make informed decisions, allocate resources effectively and minimise the social and economic impact of disinformation.

With the development of modern technologies and growing access to information, the amount of data related to disinformation has increased significantly. However, this data is often unwieldy, heterogeneous and distributed among numerous sources, making it difficult to manually extract useful information. A disinformation detection system must effectively collect and analyse this data, providing critical information for decision-making.

An in-depth analysis of the scientific literature was conducted to identify methods of disinformation detection, the results of which proved valuable for this study. Detecting disinformation is a complex process with a high level of inaccuracy [1, 3] and plays a key role in countering the spread of fake news and other forms of information manipulation. Modern methods of detecting disinformation need to be improved to meet the current challenges in the field of information security.

When developing the system, an important guideline was the ideas set out in a study on the use of the Passive Aggressive Classifier (PAC) for disinformation detection [2, 4-7]. PAC is an effective method for text classification, capable of identifying disinformation with high accuracy, even in the case of unstructured data or noise.

We also used the Multinomial Naive Bayes (MNB) classifier, which is one of the most common algorithms for text classification [3, 8-11]. MNB was used in conjunction with TfidfVectorizer to convert the text into a numerical format suitable for model training. TfidfVectorizer is an effective tool for converting text data into a numerical form for further processing.

As a result, an integrated model combining TfidfVectorizer, MNB classifier and PAC algorithm was created. This model demonstrated high accuracy in detecting disinformation during testing. The model was trained on a large dataset containing textual data labelled as disinformation or non-disinformation.

III. ANALOGUES OF DETECTING FALSE INFORMATION

To develop a disinformation detection system, it is worth analysing existing solutions on the market. This allows us to identify the strengths and weaknesses of existing analogues, as well as to identify unique features that can be integrated into our system. One of the well-known resources for fact-checking and identifying false information is Fact-checking.org (Fig. 1).

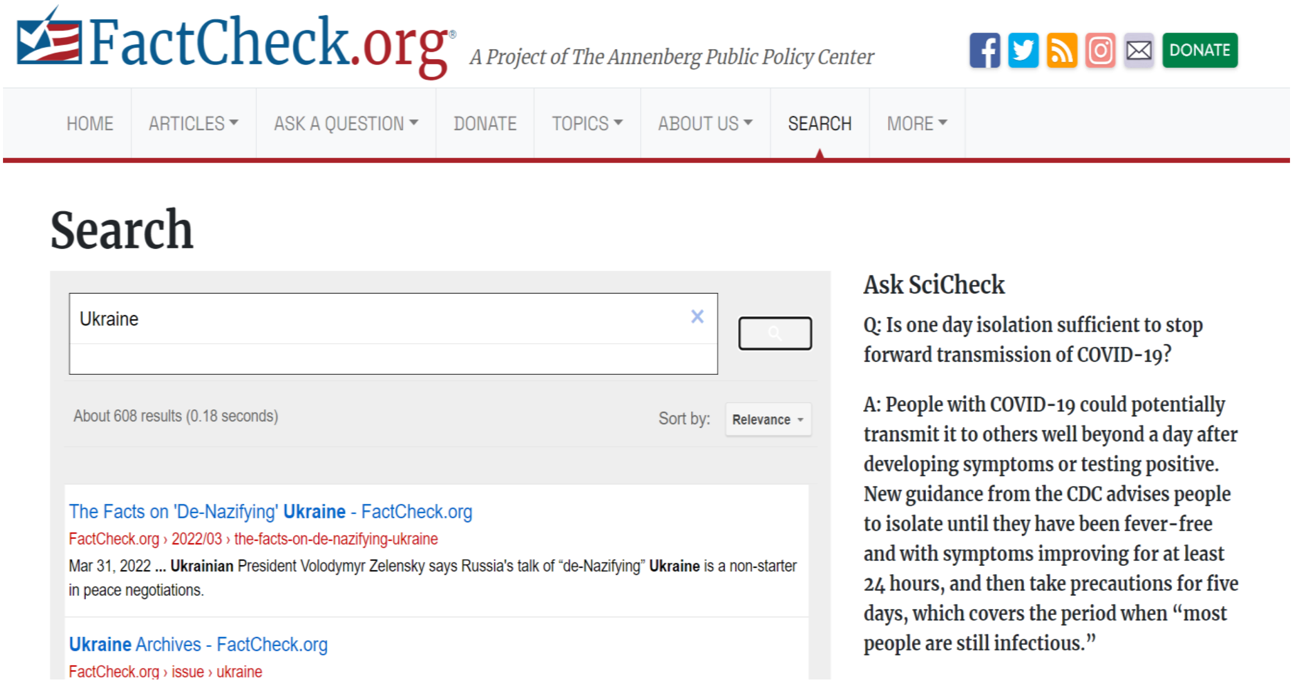
Advantages of Fact-checking.org:

- Extensive database of verified information and disinformation: Fact-checking.org has one of the largest databases, providing quick access to the information you need.
- Interactive tools for search and analysis: The platform offers user-friendly interactive tools for in-depth fact-checking and analysis.
- Specialised tools for journalists and researchers: Fact-checking.org provides additional opportunities for journalists and researchers to investigate stories in more detail and identify false information.
- Support for an international audience: The platform is accessible to users from different countries.

Disadvantages of Fact-checking.org:

- Limited options for non-English-speaking users: For those who do not speak English, the functionality of the platform may be insufficient.
- Slow database updates: Sometimes the database is updated with a delay, which can lead to the availability of outdated information.
- Limited user support: The platform may have insufficient support for new users, which can make it difficult to learn.

Another popular resource for fact-checking and identifying disinformation is Snopes (Fig. 2).



Search

Ukraine

About 608 results (0.18 seconds)

Sort by: Relevance

The Facts on 'De-Nazifying' Ukraine - FactCheck.org
FactCheck.org › 2022/03 › the-facts-on-de-nazifying-ukraine
Mar 31, 2022 ... Ukrainian President Volodymyr Zelensky says Russia's talk of "de-Nazifying" Ukraine is a non-starter in peace negotiations.

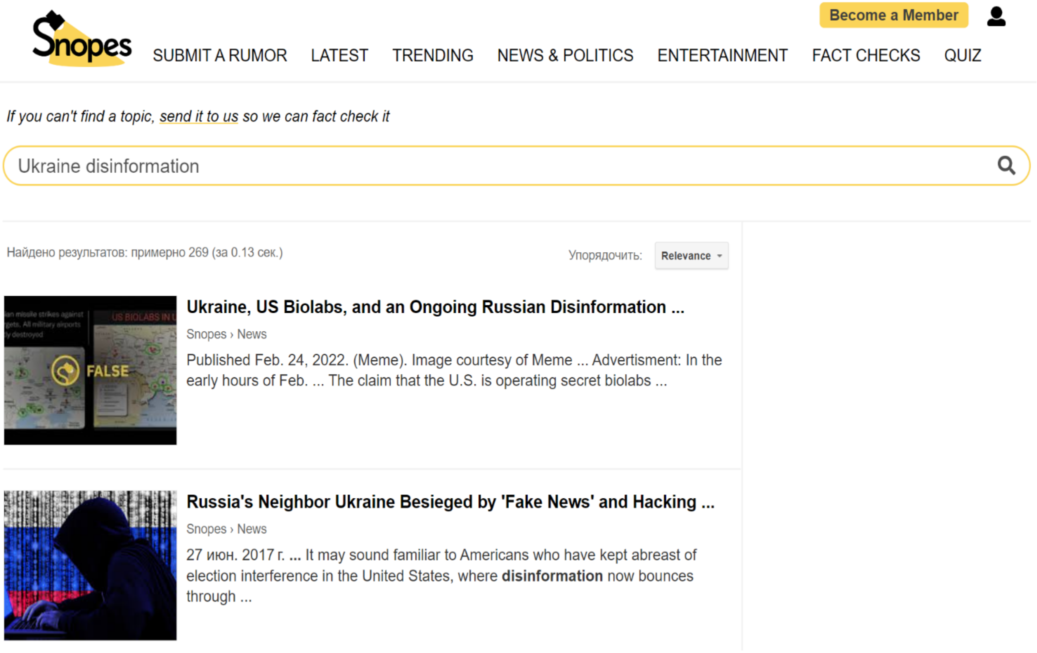
Ukraine Archives - FactCheck.org
FactCheck.org › issue › ukraine

Ask SciCheck

Q: Is one day isolation sufficient to stop forward transmission of COVID-19?

A: People with COVID-19 could potentially transmit it to others well beyond a day after developing symptoms or testing positive. New guidance from the CDC advises people to isolate until they have been fever-free and with symptoms improving for at least 24 hours, and then take precautions for five days, which covers the period when "most people are still infectious."

FIG. 1. Interface of the information search page on Fact-checking.org.



If you can't find a topic, send it to us so we can fact check it

Ukraine disinformation

Найдено результатов: примерно 269 (за 0.13 сек.)

Ukraine, US Biolabs, and an Ongoing Russian Disinformation ...
Snopes › News
Published Feb. 24, 2022. (Meme). Image courtesy of Meme ... Advertisment: In the early hours of Feb. ... The claim that the U.S. is operating secret biolabs ...

Russia's Neighbor Ukraine Besieged by 'Fake News' and Hacking ...
Snopes › News
27 июн. 2017 г. ... It may sound familiar to Americans who have kept abreast of election interference in the United States, where **disinformation** now bounces through ...

FIG. 2. Interface of the search page on Snopes.

The benefits of Snopes:

- Extensive database of verified facts and disinformation: Snopes has one of the largest databases, allowing users to quickly find the information they need.

- Specialised tools for search and analysis: The platform offers useful tools for detailed search and analysis of facts, allowing users to conduct in-depth analysis of information.

- Social media support: Snopes provides support through social media, which allows users to keep up to date with updates and news about the platform.

- Free access: Snopes offers a free version that allows users to use the platform for free.

Disadvantages of Snopes:

- Limited options for non-English-speaking users. For those who do not speak English, the platform's functionality may be insufficient.

- Slow database updates. Sometimes the Snopes database is updated with a delay, which can lead to outdated information.

- Limited support for users. Snopes may have insufficient support for new users, making it difficult to learn the platform.

The development of disinformation detection software is becoming increasingly important in today's information environment, where the volume and complexity of disinformation data is growing rapidly. This underscores the need to use advanced technologies and intelligent algorithms to obtain reliable information and make informed decisions.

IV. PURPOSE AND OBJECTIVES OF THE STUDY

The goal of the study is a system that effectively detects disinformation. To achieve this goal, the following tasks have been identified:

1. Collect and combine disinformation data from various sources.

2. Preparing the collected data by cleaning and converting it into a format suitable for analysis, using machine learning methods to draw meaningful conclusions.

3. Developing and tuning disinformation detection models to predict future patterns based on historical data and relevant characteristics, testing different algorithms to determine the most accurate approach.

4. Develop the server side of the application using Node.js (Express.js) to process requests and interact with the MongoDB database.

5. Creation of APIs for interaction with the frontend and integration of disinformation detection models for data processing and search.

6. Creating a user-friendly interface using React.js, Redux Toolkit, and Tailwind CSS to quickly create styles and components.

7. Implementation of Axios to communicate with the server API and exchange data between the client and server via HTTP requests.

8. Conducting comprehensive testing to ensure the accuracy, reliability and stability of the software, with unit, integration and system tests to identify and resolve any issues or bugs.

V. MODELS AND METHODS

To analyse and classify news articles, various statistical models and indicators are used to help identify patterns and relationships in the data. The dynamics of news is studied through a system of indicators that reveal various aspects of this process. The main sources of information are official news articles, and additional sample studies are aimed at identifying the reasons for the spread of fake news:

1. Text vectorisation (TF-IDF) is an important stage of natural language processing (NLP) and information retrieval. Its goal is to convert unstructured text data into a numerical format that can be used by machine learning algorithms.

The advantages and disadvantages of text mining methods are shown in Table 1.

TABLE 1. Advantages and disadvantages of text mining methods.

Method	Advantages	Disadvantages
Natural language processing (NLP)	Ability to understand the content and context of the text	1. Limited accuracy when studying large data sets. 2. Dependence on the amount of information.
Machine learning and analytics	Highly capable of pattern recognition and adaptation to conditions	1. The need for a large data set for training. 2. The difficulty of working with unstructured material. 3. Dependence on the amount of information.

One of the most common vectorisation methods is TF-IDF term frequency - the inverse of document frequency.

$$TF-IDF(t,d) = TF\ (t,d) \times IDF(t), \qquad (1)$$

where $TF\ (t,d)$ denotes the frequency of term t in document d, and $IDF(t)$ is the inverse document frequency for term t. TF-IDF takes into account the significance of each term in the document and its rarity in the overall corpus.

2. News classification MNB is a popular machine learning algorithm for text classification. It works on the basis of Bayes' theorem, which determines the probability of a certain event given prior knowledge of the conditions that may be associated with this event.

$$P(y|X) = P(y) \times \Pi\ (P(x\_i|y)\ /\ P(x\_i))\ , \qquad (2)$$

where $P(y|X)$ is the a posteriori probability of class y given the feature vector X; $P(y)$ is the a priori probability of class y; $P(x\_i|y$ is the probability of feature x_i for class y; and $P(x\_i)$ is the a priori probability of feature x_i.

Multinomial Naive Bayes Classifier (MNBC). It is a type of naive Bayesian algorithm used to classify text data [12-15], especially in document analysis and news classification tasks, where features are represented as frequency values (for example, the number of times a term appears in a document). MNBC applies the Bayes formula to calculate the probability of a text belonging to a certain class, assuming that the features are independent, i.e. that each feature (word) is independent of other features within the same class. An important stage of data preparation for this algorithm is text vectorisation MNBC. It is a type of naive Bayesian algorithm used to classify text data [16-18], especially in document analysis and news classification tasks, where features are represented as frequency values (for example, the number of times a term appears in a document). MNBC applies the Bayes formula to calculate the probability of a text belonging to a certain class, assuming that the features are independent, i.e. that each feature (word) is independent of other features within the same class. An important stage of data preparation for this algorithm is text vectorisation, for example, using TF-IDF, which converts text into a numerical format, allowing the algorithm to work with text data efficiently., for example, using TF-IDF, which converts text into a numerical format, allowing the algorithm to work with text data efficiently. The MNBC algorithm is based on the assumption of feature independence, which simplifies the calculation of the a posteriori probability. However, this assumption is not always accurate, and more sophisticated algorithms may be required to more accurately reflect the relationships between features.

PAC is an online learning algorithm used for classification tasks [19-20]. Its principle is to minimise the loss function at each step while controlling the level of aggressiveness during model updates.

$$\mathrm{L}(\mathrm{y},\mathrm{y}') = \max(0, 1 - \mathrm{y} \times \mathrm{y}') \qquad (3)$$

where $\mathrm{L}(y,y')$ is the loss function, y is the real class, and y' is the predicted class.

The PAC algorithm is known for its ability to adapt to changes in the data distribution and its robustness to anomalies. At the same time, it may perform poorly on datasets with a large number of dimensions or in the presence of noise.

3. Model evaluation (classification accuracy). Accuracy is a popular metric for evaluating classification models that measures the percentage of correctly classified samples among all samples in a test set.

$$\mathrm{Accuracy} = (\mathrm{TP} + \mathrm{TN})\ /\ (\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}), \qquad (4)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Precision scores provide a simple and intuitive approach to evaluate the performance of a classification model, but may be less useful for unbalanced datasets where one class is significantly overrepresented.

4. The confusion matrix is a more detailed evaluation metric that provides a comprehensive picture of the classification model's performance. It is a table used to check the model's performance on test data.

$$\mathrm{CM} = [[\mathrm{TP}, \mathrm{FP}], [\mathrm{FN}, \mathrm{TN}]], \qquad (5)$$

where CM is the confusion matrix, TP is the number of true positives, FP is the number of false positives, FN is the number of false negatives, and TN is the number of true negatives. The confusion matrix provides a wealth of information about the performance of the classification model, including accuracy, precision, recall, and F1 score. It is a useful tool for identifying model strengths and weaknesses and for comparing the performance of different models.

The development of logic, design and functionality is a key stage in the creation of an intelligent system for disinformation detection. This stage includes forming the structure of the system, identifying its main components, their interaction and developing the logic of the system.

The basic logic is to analyse the text entered, detect disinformation using machine learning algorithms and present the results to the user via a web interface. The system logic can be divided into the following stages:

1. Data collection: obtaining textual data from various sources (news articles, blogs, etc.).

2. Data processing: textual data is processed by tokenisation, lemmatisation and stop word removal. Data sets are selected and formed to train a machine learning model.

3. Model training: application of machine learning algorithms (such as Naive Bayes, Support Vector Machines, Random Forest) to train models based on text data. The models are evaluated and tuned to achieve high accuracy in detecting disinformation.

4. Disinformation detection: using trained models to identify disinformation in texts.

5. Presentation of results: development of a web interface to visualise the results for users.

VI. RESULTS OF MODEL IMPLEMENTATION

Let us describe the approach to developing an intelligent system for detecting disinformation The first step was to collect relevant data from reliable sources containing news articles marked FAKE or REAL (Fig. 3).

It is compiled from a variety of reliable sources, including official news resources and relevant data repositories. A thorough data collection process ensures that the dataset is based on reliable and up-to-date information, providing a comprehensive and accurate reflection of the latest articles.

After collection, the data underwent a detailed process of cleaning, integration and transformation to ensure that it is suitable for analysis and accurate identification of classified information. Several stages of pre-processing were carried out:

1. Identification and processing of missing values: Missing data can create difficulties during modelling and lead to inaccurate results. The .isnull().any() method was used to check for missing values, which returns a boolean value indicating whether each column contains missing values.

2. The next step was to split our dataset into training and test sets. This is an important step in machine learning as it allows us to evaluate the performance of our model on unknown data. We used the train_test_split function from sklearn.model_selection to split the data into training and test sets in an 80/20 ratio. This split allowed us to train the model on the training data and evaluate its performance on the test data.

3. The last stage of data preparation was the vectorisation of text data. Text data cannot be directly used in machine learning models, so it needs to be converted into a numerical form. For this purpose, we used the TfidfVectorizer class from sklearn.feature_extraction.text.

These preprocessing steps prepared the dataset for further modelling and evaluation. Checking for missing values, splitting the data into training and test sets, and vectorising the data into a numerical form all contributed to improving the overall quality and reliability of the disinformation detection model.

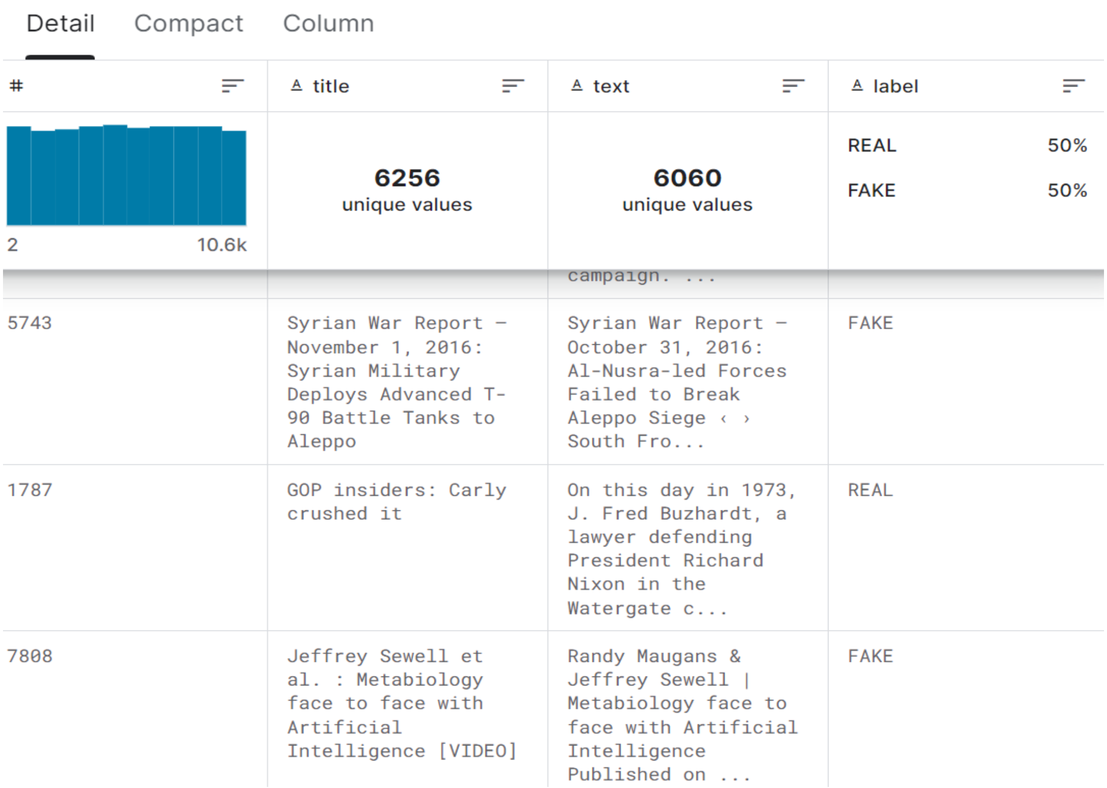
To ensure effective work with unstructured data, thanks to the flexibility and high processing speed for the intelligent disinformation detection system, a database (Fig. 4) based on MongoDB was created. It made it possible to scale the system to work with large amounts of information and guarantee fast access to data.

The database of information threat risk management in IT using intelligent disinformation detection systems contains the following collections:

1. Users stores data about registered users, including their profiles and settings;

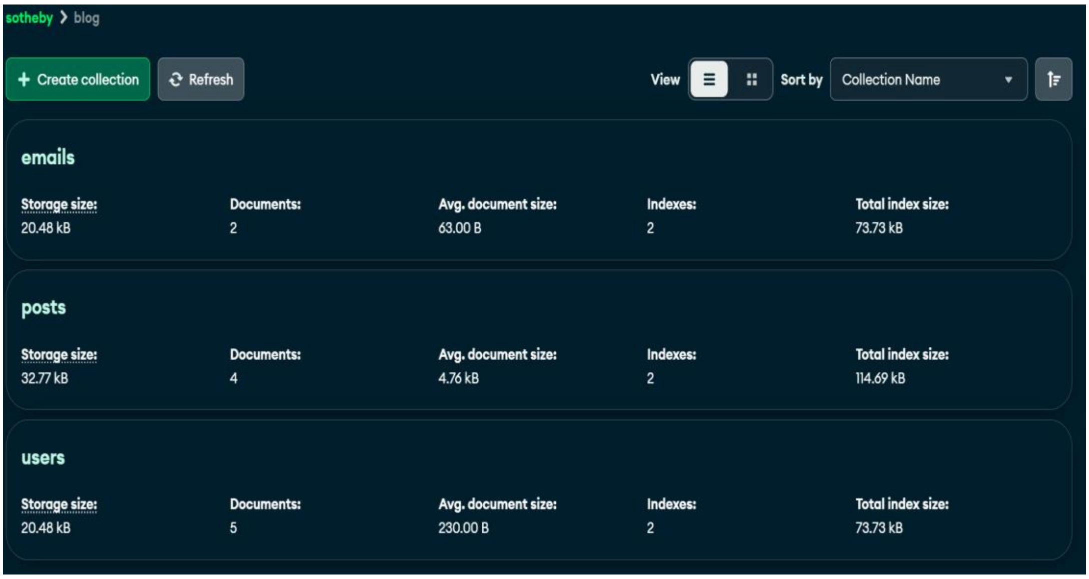
2. Posts – contains information about user publications, including text, date of creation and authors.

3. Emails stores data about email accounts.



#	title	text	label
	6256 unique values	6060 unique values	REAL 50% FAKE 50%
		campaign. ...	
5743	Syrian War Report – November 1, 2016: Syrian Military Deploys Advanced T-90 Battle Tanks to Aleppo	Syrian War Report – October 31, 2016: Al-Nusra-led Forces Failed to Break Aleppo Siege ‹ › South Fro...	FAKE
1787	GOP insiders: Carly crushed it	On this day in 1973, J. Fred Buzhardt, a lawyer defending President Richard Nixon in the Watergate c...	REAL
7808	Jeffrey Sewell et al. : Metabiology face to face with Artificial Intelligence [VIDEO]	Randy Maugans & Jeffrey Sewell | Metabiology face to face with Artificial Intelligence Published on ...	FAKE

FIG. 3. Structure of the dataset for disinformation detection.



emails

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
20.48 kB	2	63.00 B	2	73.73 kB

posts

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
32.77 kB	4	4.76 kB	2	114.69 kB

users

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
20.48 kB	5	230.00 B	2	73.73 kB

FIG. 4. MongoDB Compass database collections.

For convenient data management, the study chose the Mongoose object-document modelling (ODM) tool. It allows you to define the structure of documents, set validation rules, and greatly simplifies the work with the database. In addition, Mongoose has built-in mechanisms for verifying data before it is saved.

To interact with the database, Node.js is used in conjunction with Express.js, which ensures efficient query processing and data interaction. Thanks to Express.js, you can easily create RESTful APIs, which contributes to high performance and scalability of the system.

Training a disinformation detection model is a critical step in analysing textual data, as it determines the accuracy and consistency with which the model distinguishes between fake and true news. The training process consists of several key stages, each of which is aimed at improving the model's performance and adapting it to real-world conditions.

The first step is to convert textual data (news articles) into numerical vectors using TF-IDF (term frequency inverse of document frequency). TF-IDF takes into account the frequency and importance of each word in the text in the context of the entire dataset. This method allows the model to better understand the semantics of the text and identify key elements that distinguish reliable news from false news. For example, terms that are more frequent in certain news categories may have a high TF-IDF, which increases their importance for classification.

After text vectorisation, it is important to choose a model for classification. In our case, two models were chosen: Multinomial Naive Bayes and passive-aggressive classifier.

Multinomial Naive Bayes is a popular text classification algorithm due to its efficiency and ease of working with large amounts of data. It is based on the probability that each word in a document belongs to a certain category.

The passive-aggressive classifier is suitable for online learning tasks that require the model to constantly update its knowledge and quickly adapt to changing conditions.

After selecting a classification model, it is trained on training data. During the training process, the model adjusts its internal parameters according to the data attributes, which allows it to learn to distinguish between classes (fake and real news) based on TF-IDF vectors.

Once training is complete, the next step is to evaluate the model (Fig. 5) to determine how well it classifies news. For this purpose, various indicators are used to better understand the model's performance.

The first metric we look at is accuracy, which shows the percentage of news items correctly identified by the model. In this case, the accuracy is 0.85, meaning that the model successfully classified 85% of all news, which gives a general idea of its effectiveness.

Next, let's look at the recall metric, which reflects the proportion of true news correctly identified by the model among all real news in the test set. The algorithm identified 98% of the true news, which is a very high rate. However, only 71% of the false news was correctly identified, which indicates that there is a need for improvement in fake news detection.

The third indicator is accuracy, which indicates the percentage of news correctly classified as fake or true. The model is highly accurate in classifying fake news (98%), but less accurate in classifying true news (77%), which is important to consider when evaluating the results.

Finally, the F1 score (Fig. 6) is a harmonic average of accuracy and recall. For fake news, the F1 score is 0.82, and for true news, it is 0.87, indicating that the model demonstrates a balanced combination of accuracy and recall for both categories.

For a deeper analysis of the model's performance, the confusion matrix was used (Fig. 7). It showed that the model correctly identified 443 fake news stories and 628 true ones. However, there were still some errors: 185 fake news items were mistakenly classified as true, and 11 true news items were classified as fake.

In addition, we analysed the ROC-AUC (Fig. 8), which demonstrates how effectively the model distinguishes fake news from true news. The model achieved a high ROC-AUC value of 0.9677, which indicates its high ability to recognise both classes.

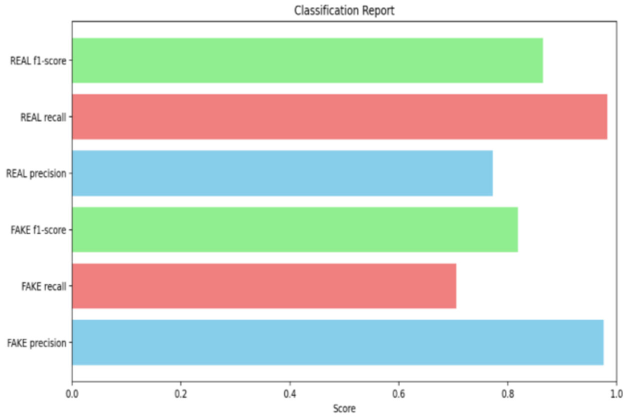
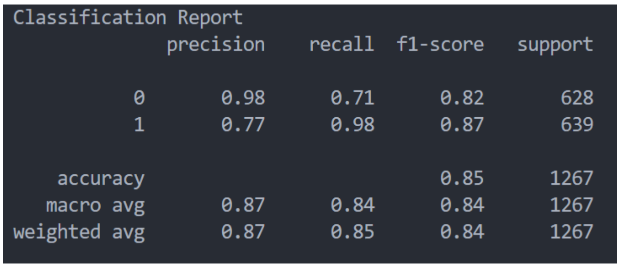


FIG. 5. Performance analysis of a trained model for news classification.



```
Classification Report
              precision    recall  f1-score   support

           0       0.98      0.71      0.82       628
           1       0.77      0.98      0.87       639

    accuracy                           0.85      1267
   macro avg       0.87      0.84      0.84      1267
weighted avg       0.87      0.85      0.84      1267
```

FIG. 6. Visualisation of model results.

The user can use the main tool of the system (Fig. 9) - the disinformation search system - to assess the probability of the reliability of the selected resource.

After entering the link, the system automatically analyzes the content of the article, using machine learning and natural language processing algorithms to identify possible signs of disinformation.

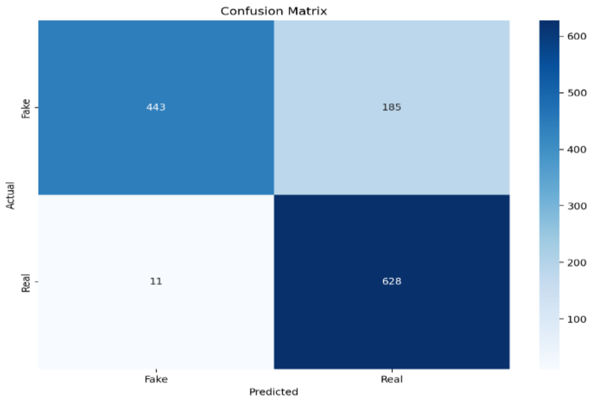


FIG. 7. Demonstration of the confusion matrix.

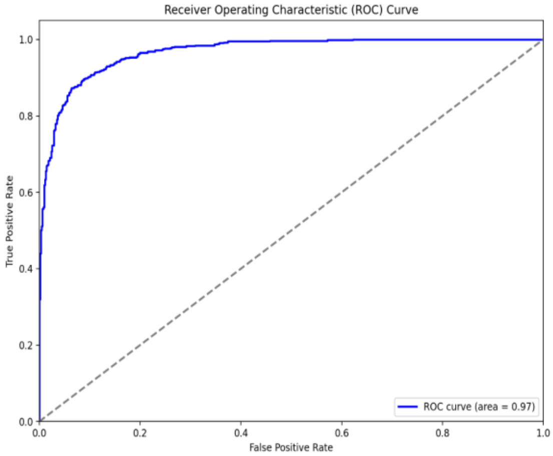
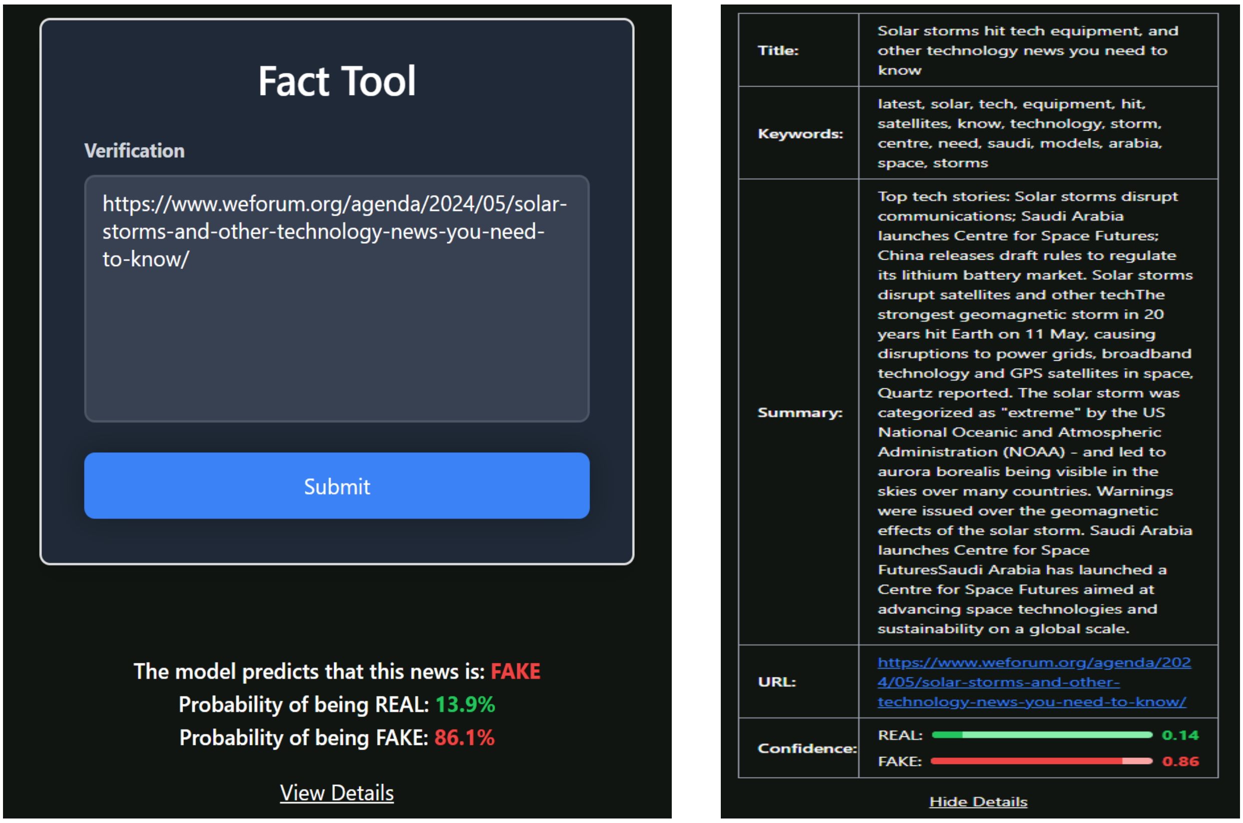


FIG. 8. Performance analysis of a trained model for news classification.



Fact Tool

Verification

https://www.weforum.org/agenda/2024/05/solar-storms-and-other-technology-news-you-need-to-know/

The model predicts that this news is: FAKE
Probability of being REAL: 13.9%
Probability of being FAKE: 86.1%

Title:	Solar storms hit tech equipment, and other technology news you need to know
Keywords:	latest, solar, tech, equipment, hit, satellites, know, technology, storm, centre, need, saudi, models, arabia, space, storms
Summary:	Top tech stories: Solar storms disrupt communications; Saudi Arabia launches Centre for Space Futures; China releases draft rules to regulate its lithium battery market. Solar storms disrupt satellites and other techThe strongest geomagnetic storm in 20 years hit Earth on 11 May, causing disruptions to power grids, broadband technology and GPS satellites in space, Quartz reported. The solar storm was categorized as "extreme" by the US National Oceanic and Atmospheric Administration (NOAA) - and led to aurora borealis being visible in the skies over many countries. Warnings were issued over the geomagnetic effects of the solar storm. Saudi Arabia launches Centre for Space FuturesSaudi Arabia has launched a Centre for Space Futures aimed at advancing space technologies and sustainability on a global scale.
URL:	https://www.weforum.org/agenda/2024/05/solar-storms-and-other-technology-news-you-need-to-know/
Confidence:	REAL: 0.14 FAKE: 0.86

FIG. 9. Interface of the system for detecting possible signs of disinformation.

VII. CONCLUSION

The developed intelligent disinformation detection system is a significant step in the fight against the spread of false information and the management of information threats. Its effectiveness is confirmed by the real results obtained during the experimental evaluation.

The following key results were obtained:

1. Methods used: TF-IDF text vectorisation, MNB classifier and PAC were used to analyse and classify news articles. The choice of these methods is based on their advantages for the task of text classification, in particular, for detecting disinformation. TF-IDF allows to extract the most significant words in texts, MNB effectively classifies texts based on frequency characteristics, and PAC provides adaptability and resistance to anomalies.

2. Data pre-processing: The collected data underwent cleaning, integration, and transformation to ensure its suitability for analysis. The methods used were identification and processing of missing values, splitting the data into training and test sets in an 80/20 ratio, and vectorisation of text data using TF-IDF.

3. Database: To efficiently work with unstructured data, a MongoDB-based database was created containing the Users, Posts, and Emails collections. Mongoose object-document modelling tool was used for data management.

4. Model development and training: Text data (news articles) were converted into numerical vectors using TF-IDF. The MNB and PAC models were used for classification.

The model was evaluated:

1. The classification accuracy is 0.85.
2. The recall rate for true news is 98%, for false news - 71%.
3. The classification accuracy for fake news is 98%, for true news - 77%.
4. The F1 score for fake news is 0.82, for true news - 0.87.
5. The confusion matrix showed that the model correctly identified 443 fake news items and 628 true news items.
6. The ROC-AUC value is 0.9677, which indicates that the model is highly capable of distinguishing between both classes.

The developed intelligent disinformation detection system is an effective tool for combating the spread of false information and managing information threats. The results of the study confirm its high accuracy and practical value. The system has significant potential for use in various fields and requires further development to adapt to the changing information environment.

VIII. FURTHER DIRECTIONS FOR RESEARCH

Despite the results achieved, research in this area needs to be continued. Further developments can be aimed at:

1. Improving classification algorithms: to improve the accuracy and completeness of disinformation detection, in particular by using deep learning and other modern methods.

2. Expanding the system's functionality: adding the ability to analyse not only textual but also visual and audio information.

3. Adaptation to new types of disinformation: developing mechanisms that will allow the system to effectively counteract new tactics of spreading fake news.

4. Integration with other systems: to create a comprehensive information security system that includes tools to detect and counter various information threats.

AUTHOR CONTRIBUTIONS

D.U., Yu.U., M.K., M.Z. – conceptualization, methodology, investigation, writing (original draft preparation), writing (review and editing).

COMPETING INTERESTS

The authors have no competing interests.

REFERENCES REFERENCES

[1] S. S. Iyengar et al., "Fake news detection using machine learning algorithms," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 9, no. 3, pp. 234-241, 2020.

[2] A. K. Singh et al., "Passive aggressive classifier for fake news detection," *Journal of Intelligent Information Systems*, vol. 57, no. 2, pp. 257-271, 2021.

[3] A. K. Mishra et al., "Multinomial naive Bayes for text classification," *Journal of Information and Communication Technology*, vol. 20, no. 2, pp. 147-162, 2020.

[4] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science*, vol. 349, no. 6245, pp. 261-266, 2015.

[5] C. C. Aggarwal and C. X. Zhai, *Mining Text Data*, Springer, 2012.

[6] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[7] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513-523, 1988.

[8] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in *Proceedings of the 15th Int. Conference on Machine Learning*, 1998, pp. 41-48.

[9] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online passive-aggressive algorithms," *Journal of Machine Learning Research*, vol. 7, pp. 551-585, 2006.

[10] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.

[11] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77-89, 1997.

[12] P. Singh et al., "A study on fake news detection using machine learning algorithms," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 9, no. 3, pp. 234-244, 2020.

[13] A. Kumar and V. Sharma, "Fake news detection using machine learning: A review," *Journal of Intelligent Information Systems*, vol. 56, no. 2, pp. 257-275, 2020.

[14] Y. Wang et al., "Deep learning for fake news detection: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 231-244, 2020.

[15] J. Thorne and A. Vlachos, "Fact-checking for fake news detection," *Journal of Data and Information Quality*, vol. 10, no. 2, pp. 1-23, 2018.

[16] F. C. Santos, "Artificial intelligence in automated detection of disinformation: A thematic analysis," *Media Journal*, vol. 4, no. 2, pp. 679-687, 2023.

[17] J. Osborne and E. Briant, "AI-driven tools for risk management in IT security and misinformation detection," *Cyber Security Review*, vol. 6, no. 2, pp. 120-133, 2021.

[18] R. Khan and A. Shaikh, "Detection of fake news and its implications on IT risk management," *International Journal of Digital Information*, vol. 9, no. 3, pp. 210-225, 2022.

[19] P. Dahlgren, "Risk mitigation strategies through AI-assisted disinformation detection," *Journal of Information Security Research*, vol. 12, no. 5, pp. 340-354, 2021.

[20] K. Jain and T. Verma, "Leveraging AI for proactive risk management against misinformation threats in IT infrastructure," *Cyber Intelligence Review*, vol. 9, no. 2, pp. 70-84, 2024.

Dmytro Uhryn

Graduated from Yuriy Fedkovych Chernivtsi National University, Chernivtsi. He is currently a Doctor of Technical Sciences, Professor, Associate Professor at Yuriy Fedkovych Chernivtsi National University. Research interests: data mining, decision support information technologies, swarm intelligence systems, industry-specific geographic information systems.

ORCID ID: 0000-0003-4858-4511

Myroslav Kovalchuk

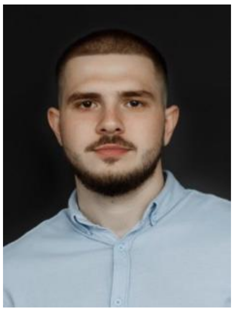
Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Computer Science, Yuriy Fedkovych Chernivtsi National University. His research interests include neural networks, the development of information systems, and the organization of databases.

ORCID ID: 0000-0001-8218-3493

Yuriy Ushenko

Prof., Computer Science Department, Chernivtsi National University, Chernivtsi, Ukraine. Research Interests: Data Mining and Analysis, Computer Vision and Pattern Recognition, Optics & Photonics, Biophysics.

ORCID ID: 0000-0003-1767-1882



Mykyta Zakharov

PhD student, Computer Science Department, Chernivtsi National University, Chernivtsi, Ukraine. Has publications in student scientific conferences. Research Interests: Data Mining, Artificial Intelligence and Analysis

ORCID ID: 0009-0003-5026-3546

Управління ризиками інформаційних загроз в ІТ за допомогою інтелектуальних систем виявлення дезінформації

Дмитро Угрин[1,*], Юрій Ушенко[1], Мирослав Ковальчук[1], Микита Захаров[1]

[1]Кафедра комп’ютерних наук/Відділ комп’ютерних технологій/Навчально-науковий інститут фізико-технічних та комп’ютерних наук, Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна

* Автор-кореспондент (Електронна адреса: d.ugryn@chnu.edu.ua)

АНОТАЦІЯ У статті описано інтелектуальну систему виявлення дезінформації, що протидіє поширенню недостовірної інформації в інтернеті. Використно сучасні методи обробки тексту і даних, зокрема машинне навчання та обробку природної мови, для точної ідентифікації фейкових новин. Основною функцією системи є надання прогнозованої оцінки достовірності інформації, що допомагає користувачам приймати обґрунтовані рішення, мінімізуючи ризик потрапляння під вплив дезінформації. Тестування системи в реальних умовах підтвердило її здатність швидко ідентифікувати фейкову інформацію, сприяючи інформаційній грамотності користувачів та підвищенню їх обізнаності щодо інформаційних загроз. Така система є важливим елементом посилення інформаційної безпеки, що особливо актуально для України в умовах численних інформаційних викликів. Система також відіграє важливу роль у керуванні ризиками інформаційних загроз в ІТ. Результати дослідження дали змогу виявляти потенційні загрози у вигляді дезінформації, які можуть бути використані для маніпуляцій громадською думкою чи підриву довіри до інституцій. Інтеграція інтелектуальних систем у процеси управління ризиками дозволяє вчасно реагувати на загрози, знижуючи їхній вплив на ІТ-інфраструктуру та зберігаючи репутацію організацій. Система може бути застосована не тільки в ІТ-секторі, а й у журналістиці, освіті, державному управлінні, де вона допомагає запобігати дезінформації, що має серйозний соціальний вплив. Її також можна використовувати для моніторингу та аналізу інформаційних потоків, що сприяє виявленню і протидії неправдивій інформації. Таким чином, розроблена система є важливим кроком у посиленні інформаційної безпеки, забезпечуючи захист від фейкових новин та слугуючи ефективним інструментом управління ризиками інформаційних загроз у сучасному цифровому суспільстві.

КЛЮЧОВІ СЛОВА інтелектуальна система, виявлення дезінформації, штучний інтелект, машинне навчання, глибинне навчання.