UDK 004.032.26:070.16 **DOI:** 10.31673/2412-9070.2024.045257

M. S. HNATYSHYN, PhD student,
ORCID: 0009-0009-0813-3602;
O. L. NEDASHKIVSKIY, D. S., assoc. professor,
ORCID: 0000-0002-1788-4434,
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv

METHODS AND SOFTWARE FOR THE DETECTION OF FALSE INFORMATION ON THE INTERNET BASED ON NEURAL NETWORKS

The article contains an overview of modern methods and software based on neural networks for detecting false information on the Internet, as well as an analysis of current problems and possible directions of future research in this field. The results of the work can be used as an information base for continuing research in the direction of using neural networks of various types to prevent the spread of fake information on the Internet.

With the dawn of the digital age and the popularity of online social networks, information is spread faster and easier than ever before. However, it also promotes the spread of poor quality or intentionally fake information, which can have a negative impact on society. Identifying, flagging and refuting disinformation on the Internet as quickly as possible is becoming an increasingly urgent problem.

The article shows that modern methods of using neural networks in detecting false information are highly effective due to their ability to process large volumes of data and detect complex patterns. The use of graph neural networks, behavioral analysis and other innovative technologies provides wide opportunities for adapting detection systems to different requirements and conditions, which allows developing more flexible and effective solutions that can work in different contexts and with different types of data.

An important advantage of neural networks and their software implementation is the possibility of integration of various data sources and contextual information. This allows software information systems not only to analyze the textual content of news, but also to take into account social interactions, the history of publications and other factors that may indicate the falsity of information.

It is shown that a very important aspect of false information on the Internet is images and videos that are presented with a false interpretation or with modification. The ability of the future system to recognize such cases will significantly increase the effectiveness of determining the reliability of information.

But despite the advantages, there are problems such as high computational requirements and difficulties in interpreting the results. This requires further research and improvements, especially in the area of improving algorithms and developing more efficient and scalable solutions and software. In addition, an important direction is the development of methods for early detection of fake news and minimizing their impact on public opinion.

Keywords: social networks; fake news; false information; neural networks; software engineering.

Introduction

With the advent of the digital age and the popularity of online social networks, information is spread quickly and easily like never before. However, it also promotes the dissemination of low-quality or intentionally fake information, which can have a negative impact on society. Identifying, flagging, and disproving online misinformation as quickly as possible is becoming an increasingly pressing issue.

Fact checking as a journalistic tool has grown significantly from 11 sites in 2008 to 424 in 2022. However, since 2019, there has been a decline in the number of new fact-checking sites. In 2023, there were 417 active fact-checking sites operating in more than 100 countries and 69 languages. This indicates a stabilisation in the development of fact-checking, despite the growing need to combat manipulative media and political lies.

The problem of the ability to distinguish genuine news from fiction is relevant for our society. According to a sociological survey by the InMind company, only 68% of Ukrainian citizens can confidently distinguish disinformation using available tools (fig. 1) [9]. Therefore, new tools should increase the share of such people.

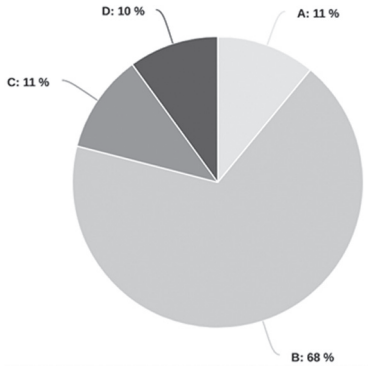

Fig. 1. Results of a sociological experiment on the definition of disinformation: A — all 3 answers are correct; B — at least 1 answer is correct; C — all answers are not correct; D — didn't answer all questions

On the other hand, the role and effectiveness of deep learning and graph neural networks in detecting fake news is growing. These methods allow us to effectively model the social context and process of online news dissemination. Currently, there are three main approaches to using graph neural networks: knowledge-based methods, propagation-based methods, and heterogeneous social context-based methods.

Knowledge-based methods use elements of news content, such as concepts and named entities, to identify fake news. Such methods can use both internal and external sources of knowledge to build a more complete picture of the reliability of information. Dissemination-based methods focus on the process of news distribution and user interaction with that news. Methods based on heterogeneous social context include the broader context of the news source, such as other posts by the same user or news on similar topics.

This article provides an overview of current techniques and software based on neural networks for detecting false information on the Internet and discusses current challenges and possible directions for future research in this area.

Analysis of literature and statement of the problem

Currently, there are many scientific works in the field of detecting and combating disinformation. They can be divided into several directions: methods of determining the veracity of textual information, methods of determining the authenticity and originality of graphic information, and methods of finding networks of fake users that are created to influence society for one purpose or another. Very recently, scientific works related to the detection of informational and psychological special operations, which is a very urgent problem in our country, have been added to this list.

Information and psychological special operation is a type of activity aimed at influencing public opinion, emotions, psychological state or behavior of a certain group of people or the population as a whole with the help of a specially organized information campaign. This activity may include the use of mass media, social networks, propaganda, disinformation, psychological pressure and other methods of information influence. The main purpose of such operations is to achieve specific political, social or military goals by manipulating information and forming a certain perception of events or facts in the target audience.

In the field of detecting fake news and bots on the Internet, current research offers a wide range of methods and approaches based on neural networks [3]. Considering various studies, it is possible to distinguish key directions:

1. Graph Neural Networks in Fake News Detection: the main focus of this field is the development of graph models that enable the detection of fake news using a variety of data sources and contextual information. Such methods allow comprehensive analysis of social interactions related to the dissemination of information, including knowledge-oriented, dissemination-based, and heterogeneous social context methods [6]. They provide a deep understanding of the dynamics of social networks, but face problems related to high computational requirements and the complexity of interpreting the results;

2. Bot detection based on text and behavior analysis: other studies focus on detecting bots in social networks using text analysis techniques and behavioral models [1]. Detecting bots is a key aspect of fighting fake news, as bots are often used to manipulate public opinion and spread misinformation [8].

Most of the works focus only on information in social networks, while bypassing other sources of information, such as: news sites, personal journalistic blogs, government sites, etc. To build more effective methods of detecting disinformation, it is necessary to include all possible sources, this will allow a better assessment of the general context and a more balanced decision regarding this or that news.

Based on the literature review, it can be noted that combining different disinformation detection methods and bots may involve the use of different deep learning models that analyze text, behavioral features, metadata, and network structure. Such a comprehensive approach can improve the accuracy and ability to recognize more sophisticated and evolved forms of bots and disinformation on social networks.

The challenge that can be solved by combining these approaches involves developing a more flexible and accurate detection system that can adapt to changes in bot behavior and disinformation dissemination methods.

Purpose and objectives of research

The purpose of the research is to develop and improve methods and software based on neural networks for detecting false information in real time. This involves developing algorithms that can detect complex patterns of misinformation and adapt to constant changes in the strategies and methods of spreading false information online. The result of such research will be the creation of effective tools designed to increase the reliability and level of security of the information space in the network.

The research objectives include the creation of a new model for detecting false information on the Internet using the method of neural networks, by improving existing algorithms that take into account the complexity and dynamics of modern disinforma-

tion dissemination strategies. Methods that can take into account a wide range of parameters, including structural and semantic features of information flows, which will make them more effective and universal compared to traditional approaches.

The study is designed to increase the level of information security on the Internet for society as a whole and for ordinary users, including. As part of this, it is proposed to create open access to the developed software by creating a website that will allow any user to check this or that information from social networks, news sites, etc.

An important aspect is not only providing an assessment of the reliability of information, but also substantiating such an assessment in a form that will be understandable to an ordinary person. We have in mind a deep analysis of all aspects of disinformation in one or another data, with the selection of specific cases and the provision of a detailed report as a result of the implementation of the program.

A special focus should be on graphic content, which often accompanies information on social networks or news sites. Very often, manipulation of the audience takes place precisely thanks to photo or video materials, to which a completely different description and shooting location is added than is actually the case.

Categorization of fake news

In an era marked by the rapid dissemination of information through digital platforms, the proliferation of fake news has emerged as a significant challenge. The deliberate dissemination of false or misleading information with the intent to deceive poses threats to various aspects of society, including public discourse, democratic processes, and individual decision-making. Detecting and combating fake news require comprehensive approaches, among which categorization frameworks play a crucial role [5]. Categorization enables the classification of diverse forms of misinformation, facilitating targeted detection strategies and intervention efforts.

Taxonomy, as applied to fake news, involves the systematic classification of misinformation based on various criteria such as content, intent, and dissemination mechanisms [2]. One commonly employed taxonomy distinguishes between different types of fake news based on their content, including fabricated content, misleading content, and manipulated content. Fabricated content entails entirely false information fabricated to deceive, while misleading content involves the distortion or manipulation of genuine information to mislead readers. Manipulated content refers to media, such as images or videos, that have been altered to convey a false narrative.

Another taxonomy considers the intent behind the creation and dissemination of fake news, categorizing it into categories such as political propaganda, financial fraud, and malicious hoaxes [7]. Political propaganda aims to influence public opinion or sway electoral outcomes by disseminating biased or false information.

Furthermore, taxonomies may consider the dissemination mechanisms employed by fake news, distinguishing between organic dissemination and coordinated campaigns. Organic dissemination refers to the spread of misinformation through individual users sharing content on social media platforms or other online channels. In contrast, coordinated campaigns involve organized efforts by malicious actors, such as bots or troll farms, to amplify fake news and manipulate online discourse.

In addition to taxonomies, dimensional frameworks offer another approach to categorizing fake news by considering multiple dimensions or attributes simultaneously. One such framework considers the credibility, novelty, and emotional appeal of fake news content. Emotional appeal refers to the use of emotional language or imagery to evoke strong emotional reactions from the audience, which can enhance the spread and impact of fake news.

Another dimensional framework considers the virality, veracity, and virulence of fake news. Virality refers to the speed and extent of the spread of misinformation across online networks, with highly viral fake news often reaching a large audience rapidly. Veracity refers to the truthfulness or accuracy of the information, with fake news being characterized by falsehoods or distortions of reality. Virulence refers to the potential harm or negative consequences associated with the spread of fake news, including social polarization, erosion of trust, and real-world impacts such as violence or public health crises.

Summarizing existing works, we can definitely outline at least 5 categories of the fake news that have more or less clear boundaries: false/deceptive, misleading, slanted/ biased, manipulated and humor information. The detailed description and properties of each category can be found in the table 1.

Categorization frameworks play a critical role in the detection and mitigation of fake news by enabling researchers and practitioners to classify and analyze diverse forms of misinformation systematically. Taxonomies provide a structured approach to categorizing fake news based on content, intent, and dissemination mechanisms, while dimensional frameworks offer a multidimensional perspective by considering attributes such as credibility, novelty, and emotional appeal. By leveraging these frameworks, stakeholders can develop more effective strategies for identifying, debunking, and combating fake news, thereby safeguarding the integrity of information ecosystems and promoting informed decision-making in society.

Table 1

Description of main fake news categories

Category	Description
False/deceptive	• Fictional narratives with no basis in reality. • Intentionally concocted news aimed at deceiving or generating revenue from online engagement. • Websites entirely devoted to spreading false information, often mimicking legitimate sources. • This falls under the conventional definition of fake news.
Misleading	• Narratives lacking verifiable truth yet advocating a particular agenda. • Such narratives frequently distort a fragment of factual content to suit their own narrative. • Instances falling within this classification aim to provoke emotional reactions.
Slanted/biased	• Narratives incorporating elements of truth while selectively presenting or excluding facts to advance an agenda, often for the sake of garnering attention. • Stories in this category may not be entirely false; they convey genuine events but with a biased perspective. • Examples from news outlets such as Fox News, MSNBC, Huffington Post, and others may demonstrate this tendency (while it's not imperative to completely disregard these sources, it's important to recognize potential biases).
Manipulated	• This category encompasses modified content or imagery. • Examples would encompass manipulated or digitally altered images.
Humor (satire/parody/jokes)	• Intentionally fabricated stories devoid of harmful intent but capable of deceiving individuals. • Satirical news aims for entertainment and humor rather than deception, yet there's a risk of misinterpretation by some as genuine.

Materials and research method

A number of sources, including scientific publications, articles, and current databases, were selected for the study of methods of detecting false information on the Internet. Special attention is paid to studies that analyze the use of neural networks and their effectiveness in recognizing fake news.

The efficiency of neural networks largely depends on the amount and quality of the data on which it was trained, which is why you need to approach the selection of data sets responsibly [4]. Table 2 shows a list of existing data sets for the analysis of disinformation on the Internet, which also includes the classification of one or another component.

Table 2

Data sets for the analysis of disinformation on the Internet

Name	Subject	Data format	Number of classes
ISOT	Politics, society, busienss, sport, criminality, technologies, well-being	Text	2
Fakeddit	Society, politics	Text, image, video	2, 3, 6
LIAR	Politics	Text	6
FakeNewsNet	Society, politics	Text, image	2
Stanford Fake News	Society	Text, image, video	2
FA-KES	Politics	Text	2
BREAKING!	Society, politics	Text, image	2, 3
BuzzFeed-News	Politics	Text	4

Below is a detailed description of each of the sets:

1. SOT: Both fake news and real news from Reuters; fake news from websites that have been flagged as unreliable by PolitiFact and Wikipedia.

2. Fakeddit: An English-language multimodal fake news dataset including images, comments, and news metadata.

3. LIAR: An English-language dataset of 12,836 short policy statements collected from online broadcasts and two social networks — Twitter and Facebook — from 2007 to 2016.

4. Stanford Fake News: Fake news and satirical stories, including hyperbolic support or condemnation of an individual, conspiracy theories, racist themes, and discrediting of reliable sources.

5. FA-KES: Flagged as fake news about the conflict in Syria, such as casualties, activities, locations and dates.

6. BREAKING!: An English-language dataset created using the Stanford Fake News and BS detector datasets. The data, including news about the 2016 US presidential election, was collected from web pages.

7. BuzzFeedNews: An English-language dataset of 2,283 political articles collected from Facebook from 2016 to 2017.

8. FakeNewsNet: An English-language dataset of 422 social and political articles collected from online broadcasts and Twitter.

9. FEVER: An English-language dataset of 185,445 statements about society collected from online broadcasts.

The evaluation of the effectiveness of the selected methods is carried out by analyzing their application on real data. Validation of the obtained results,

comparison with other known methods, and analysis of potential areas for further research and improvement are included.

It is planned to use the methods of machine and deep learning to solve the problems. The first stage of the research will be the search for the most optimal combination of known approaches from these areas of computer science, with further optimization and changes to meet the needs of this research.

A combined approach is used, which includes quantitative and qualitative analysis of existing methods and software solutions. The analysis is carried out based on an assessment of the accuracy, efficiency, and adaptation capabilities of various neural networks to the conditions of the spread of misinformation on the Internet.

The research is based on the use of modern data analysis and machine learning tools, including convolutional and recurrent neural networks, graph neural networks, and artificial intelligence systems for natural language processing.

Conclusion

1. Modern methods of using neural networks in the detection of false information are highly effective due to their ability to process large amounts of data and detect complex patterns. Neural networks show significant progress in text classification and analysis, which allows detecting fake news with high accuracy. However, the effectiveness of such systems depends on the quality and volume of training data.

2. The use of graph neural networks, behavioral analysis, and other innovative technologies provides ample opportunities for adapting detection systems to different requirements and conditions. This allows for the development of more flexible and efficient solutions that can work in different contexts and with different types of data.

3. An important advantage is the ability to integrate different data sources and contextual information. This allows systems not only to analysis the textual content of news, but also to consider social interactions, the history of publications and other factors that may indicate the falsity of information.

4. A separate important aspect of false information on the Internet is images and videos that are presented with a false interpretation or with modification. The ability of the future system to recognize such cases will significantly increase the efficiency in determining the veracity of information.

5. Despite the advantages, there are challenges such as high computational requirements and difficulty in interpreting the results. This requires further research and improvement, especially in the area of improving algorithms and developing more efficient and scalable solutions. In addition, an important direction is the development of methods for early detection of fake news and minimizing their impact on public opinion.

References

1. ***Andriotis, Panagiotis, Atsuhiro Takasu.*** *Emotional Bots: Content-Based Spammer Detection on Social Media // IEEE Xplore. 1 Dec. 2018. P. 1–8 [Electronic resource]. URL:*

https://doi.org/10.1109/WIFS.2018.8630760. Accessed 12 Jan. 2024.

2. ***Fake*** *News Detection: Taxonomy and Comparative Study / Faramarz Farhangian [et al.] // Information Fusion. Vol. 103. Elsevier BV. Mar. 2024. P. 102140–40 [Electronic resource]. URL:*

https://doi.org/10.1016/j.inffus.2023.102140.

3. ***Fake*** *News Detection through Graph-Based Neural Networks: A Survey / Gong Shuzhi [et al.] // ArXiv (Cornell University), Cornell University, July 2023 [Electronic resource]. URL:*

https://doi.org/10.48550/arxiv.2307.12639.

4. ***Kudugunta Sneha, Emilio Ferrara.*** *Deep Neural Networks for Bot Detection // Information Sciences. Vol. 467. Oct. 2018. P. 312–22 [Electronic resource]. URL:*

https://doi.org/10.1016/j.ins.2018.08.019. Accessed 10 Jan. 2024.

5. ***Fake*** *News Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content / Maria D. Molina [et al.] // American Behavioral Scientist. 2019. Vol. 65, no. 2. P. 000276421987822 [Electronic resource]. URL:*

https://doi.org/10.1177/0002764219878224.

6. ***Fake*** *News Detection: A Survey of Graph Neural Network Methods / Huyen Trang Phan [et al.] // Applied Soft Computing. 2023. Vol. 139. P. 110235 [Electronic resource]. URL:*

https://doi.org/10.1016/j.asoc.2023.110235.

7. ***A Taxonomy*** *of Fake News Classification Techniques: Survey and Implementation Aspects / Dhiren Rohera [et al.] // IEEE Access. 2022. Vol. 10. P. 30367–94 [Electronic resource]. URL:*

https://doi.org/10.1109/ACCESS.2022.3159651.

8. ***Twitter*** *Bots in Cyber-Physical-Social Systems: Detection and Estimation Based on the SEIR Model / Weisha Zhang [et al.] // Security and Communication Networks. Hindawi Publishing Corporation, May 2023. P. 1–9 [Electronic resource]. URL:*

https://doi.org/10.1155/2023/6234030.

9. ***USAID-INTERNEWS.*** *Attitude of the population towards mass media and consumption of different types of media in 2019 [Electronic resource]. URL:*

https://drive.google.com/file/d/1Oi2Edvl5Srk4hS-D2KoxoKkamCarUX7f/view?pli=1. Accessed 8 Jan. 2024.

М. С. Гнатишин, О. Л. Недашківський

АНАЛІЗ МЕТОДІВ ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИЯВЛЕННЯ НЕПРАВДИВОЇ ІНФОРМАЦІЇ У МЕРЕЖІ ІНТЕРНЕТ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Стаття містить огляд сучасних методів і програмного забезпечення на основі нейронних мереж для виявлення неправдивої інформації у мережі Інтернет, а також аналіз поточних проблеми та можливих напрямків майбутніх досліджень у цій галузі. Результати роботи можуть бути використані як інформаційна база для продовження досліджень у напрямку використання нейронних мереж різних типів для запобігання поширенню фейкової інформації у мережі Інтернет.

Із початком цифрової ери та популярністю онлайн-соціальних мереж інформація поширюється швидко та легко, як ніколи раніше. Однак це також сприяє поширенню неякісної або навмисно фейкової інформації, яка може мати негативний вплив на суспільство. Якнайшвидше виявлення, позначення та спростування дезінформації у мережі Інтернет стає все більш актуальною проблемою.

Показано, що сучасні методи використання нейронних мереж у виявленні неправдивої інформації є високоефективними завдяки своїй здатності обробляти великі обсяги даних і виявляти складні закономірності. Використання графових нейронних мереж, поведінкового аналізу та інших інноваційних технологій надає широкі можливості для адаптації систем виявлення до різних вимог та умов, що дозволяє розробляти більш гнучкі та ефективні рішення, які можуть працювати в різних контекстах і з різними типами даних.

Важливою перевагою нейронних мереж та їх програмної реалізації є можливість інтеграції різних джерел даних і контекстної інформації. Це дозволяє програмним інформаційним системам не тільки аналізувати текстовий контент новин, а й враховувати соціальні взаємодії, історію публікацій та інші фактори, які можуть вказувати на неправдивість інформації.

Показано, що окремим важливим аспектом неправдивої інформації в мережі Інтернет є зображення та відео, які представлені з неправдивою інтерпретацією або з модифікацією. Здатність майбутньої системи розпізнавати такі випадки істотно підвищить ефективність визначення достовірності інформації.

Але незважаючи на переваги, існують такі проблеми, як високі вимоги до обчислень і труднощі в інтерпретації результатів. Це вимагає подальших досліджень і вдосконалень, особливо в області вдосконалення алгоритмів і розробки більш ефективних і масштабованих рішень. Крім того, важливим напрямком є розробка методів раннього виявлення фейкових новин та мінімізації їх впливу на громадську думку.

Ключові слова: соціальні мережі; фейкові новини; неправдива інформація; нейронні мережі; інженерія програмного забезпечення.