Information Technology for Detecting Fakes and Propaganda Based on Machine Learning and Sentiment Analysis

Vitalii Danylyk[1], Victoria Vysotska[1]

1 Lviv Polytechnic National University

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

This article provides a comprehensive study of modern approaches used to identify fakes and propaganda. Machine learning is emerging as a dynamic tool for pattern recognition and adaptation that facilitates real-time analysis. In addition, the article provides an analysis of propaganda based on emotional colouring, which reveals the differences between propaganda and non-propaganda. The average emotional value for propaganda news is 0.151 and for non-propaganda news is 0.116. The average degree of subjectivity for propaganda news is 0.365 and for non-propaganda news is 0.283. The average value of positive emotion for propaganda news is 0.087 and for non-propaganda news is 0.082. The average negative emotion for propaganda news is 0.064 and for non-propaganda news is 0.034. -The average value of the complex emotional colouring for propaganda news is 0.021, and for non-propaganda news - 0.010. Keywords – propaganda, fakes, NLP, natural language processing, disinformation detection, machine learning, multimodal analysis.

Vitalii Danylyk, and **Victoria Vysotska**

Lviv Polytechnic National University, Ukraine

1. Introduction

In the ever-expanding digital communications space, the dissemination of information has reached unprecedented heights, enabling the rapid spread of both accurate and misleading narratives. One of the most troubling byproducts of this digital age is the spread of fake news and propaganda, two interrelated phenomena that pose a significant threat to the integrity of information and the foundations of democratic societies.

The prevalence of fake news and propaganda has become an insidious force that has infiltrated online platforms, news outlets, and social media channels. In an age of information overload, distinguishing fact from fiction is becoming increasingly difficult for media consumers. The implications of disinformation go beyond mere confusion, as they can shape public opinion, influence the political landscape, and even fuel social discord.

The impact of fake news and propaganda is felt in a variety of areas, from political discourse and public health to social cohesion. False narratives can influence elections, undermine trust in institutions, and contribute to the polarization of communities. The consequences are not only cognitive but can also manifest themselves in real-world actions, as evidenced by cases where disinformation has fuelled civil unrest or hindered effective responses to crises.

Given the seriousness of this situation, there is an urgent need for best practices in detecting and combating disinformation. Traditional fact-checking processes struggle to keep up with the rapid spread of misleading content facilitated by digital platforms. As technology advances, so do the tactics of disinformation disseminators, requiring innovative and adaptive approaches to protect the integrity of the information ecosystem.

2. Analysis of the latest research and publications

Recent research and publications have contributed significantly to the current discourse around identifying disinformation by introducing new methodologies and improving existing approaches. This section examines key discoveries and advancements, shedding light on the current state of the industry.

Recent research has highlighted the importance of integrating multimodal analysis, combining NLP with image and video processing techniques. This holistic approach acknowledges the multimedia nature of contemporary disinformation, addressing the challenges posed by deepfakes and image-based manipulation. By simultaneously inspecting textual and visual content, researchers aim to improve the overall accuracy of detection systems.

Improvements in understandable artificial intelligence have attracted attention in recent publications. Researchers are exploring methods to make the decision-making processes of NLP and machine learning algorithms more transparent and interpretable. Not only does this increase the reliability of detection systems, but it also provides insight into complex ways to detect disinformation, contributing to a broader understanding of the field.

Recent literature highlights the growing recognition of the need for interdisciplinary collaboration. While technological solutions remain key, researchers are increasingly advocating for partnerships between technologists, sociologists, and policymakers. This collaborative approach aims to combat disinformation not only from a technological point of view but also by addressing its social and political roots.

Ethical considerations in automated detection methods have become a central theme of recent research. Scientists are exploring the potential biases built into algorithms and the ethical implications of automated content moderation. Striking a balance between effectively detecting and preserving individual freedoms is a central theme that reflects the broader societal implications of using advanced technologies to detect disinformation [1].

Recent publications highlight the challenges in assessing the effectiveness of disinformation detection systems. The development of standardized indicators and benchmarks that take into account the dynamic nature of disinformation remains a pressing issue. Researchers are actively exploring ways to create comprehensive scoring systems that take into account a variety of parameters, from false positives to the ability to adapt to evolving tactics.

Recent research and publications show a dynamic landscape characterized by innovation in multimodal analysis, a focus on understandable AI, and a growing emphasis on interdisciplinary collaboration. However, ethical considerations and difficulties in assessing system performance highlight the difficulty of addressing disinformation in an ever-evolving digital environment. These findings lay the groundwork for future developments and underscore the urgency of continuing research in this critical area.

3. The purpose of the article

The main purpose of this article is to comprehensively explore modern approaches to detecting and countering fake news and propaganda in media texts. By delving into the methodologies and technologies used in recent research, the aim is to offer a detailed understanding of the evolution of the disinformation detection landscape.

The object of the research is the most deceptive narratives – the fabric of fakes and propaganda that saturate not only the vast landscape of modern media, but also permeate the very foundations of public discourse, shaping perceptions, influencing opinions, and casting doubt on truthfulness. information in the digital age.

The subject of this study covers the practical and operational techniques used to detect fakes and propaganda, serving as a practical toolkit that researchers, technologists, and policymakers use to systematically identify and counter misleading narratives in the complex landscape of contemporary media

The scientific novelty lies in the comprehensive review and critical analysis of existing methodologies. It contributes to the synthesis of diverse perspectives, the identification of gaps or areas for improvement in current methods of detecting fakes and propaganda, as well as understanding the effectiveness and limitations of these approaches. In addition, the scientific novelty is evident in the research's ability to contextualize the current state of disinformation detection in a rapidly evolving digital media environment, providing valuable knowledge for future developments in the field.

4. Statement of the main material

Modern approaches to identifying fakes and propaganda use a combination of technological innovation, interdisciplinary collaboration, and advanced analytical techniques to navigate the complex landscape of disinformation. These approaches encompass different strategies, each of which aims to address the multifaceted nature of deceptive narratives in contemporary media [2].

Natural Language Processing (NLP). Using NLP, these approaches involve analysing language patterns in textual content. NLP algorithms analyse sentences, phrases, and common language structures to identify anomalies, inconsistencies, and patterns that indicate misinformation. Sentiment analysis and semantic analysis are integral components that help recognize the emotional tone and meaning of words.

Multimodal analysis. Acknowledging the multimedia nature of contemporary disinformation, modern approaches include multimodal analysis. It involves simultaneously analysing text, images, and videos to identify manipulations or inconsistencies in different media formats. Techniques such as image forensics and deep learning contribute to a more comprehensive understanding of deceptive narratives.

Machine learning algorithms. Advanced machine learning algorithms are trained on large datasets to recognize patterns associated with misinformation. These algorithms can adapt to new tactics used by deception providers, providing a dynamic and scalable approach to identification. In this context, supervised learning, unsupervised learning, and ensemble methods are commonly used.

By integrating these diverse strategies, modern approaches to detecting fakes and propaganda seek to stay ahead of new challenges. The combination of technological advances, interdisciplinary ideas, and a commitment to ethical practices forms a solid foundation for mitigating the impact of misleading narratives on public discourse and public welfare.

4.1. Natural Language Processing

Natural Language Processing (NLP) plays a crucial role in the field of fake detection and propaganda in textual content. NLP is a subfield of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language [3]. In the context of identifying misleading narratives, NLP is used to analyse language patterns and textual content to identify inconsistencies, misleading information, or propagandistic elements.

NLP techniques for detecting fakes and propaganda include several key aspects:

1. *Text Analysis:* NLP algorithms break down textual information by examining individual words, phrases, and the overall structure of sentences and paragraphs. This process helps identify linguistic patterns associated with misinformation.
2. *Sentiment analysis:* NLP is used to assess sentiment expressed in a piece of text. Misinformation often carries clear emotional connotations or biases, and sentiment analysis helps flag content with polarized or deceptive emotions.
3. *Semantic analysis:* Understanding the meaning of words and phrases is crucial. NLP allows systems to comprehend semantic context, allowing the identification of subtle shifts in meaning or intentional distortions.
4. *Contextual Comprehension:* NLP algorithms seek to understand the contextual nuances of language, recognizing that the meaning of words can vary depending on the surrounding text. This is especially important for detecting sarcasm, irony, or other forms of indirect communication that can be used in deceptive narratives.
5. *Named Entity Recognition (NER):* NLP techniques include NER, which involves identifying entities such as people, organizations, and places in the text. Identifying inconsistencies in the representation of objects can be a valuable clue for detecting fake information.

NLP's advantages in detecting fakes and propaganda include its scalability to process large amounts of text, real-time analysis capabilities, and its ability to adapt to evolutionary linguistic patterns. However, limitations can arise from contextual ambiguity, where the meaning of words is highly dependent on the surrounding context, and the challenges associated with updating NLP models to address new language nuances.

Natural Language Processing (NLP) offers several benefits in the context of fake news detection and propaganda, contributing to the efficiency and effectiveness of disinformation identification:

1. *Scalability.* NLP allows for automated analysis of a huge amount of textual content at scale. Given the sheer volume of information disseminated online, the scalability of NLP allows for the timely processing of large amounts of data, making it easier to identify misleading narratives across platforms.
2. *Real-time analysis.* NLP systems operate in real-time, providing the ability to quickly analyse incoming information as soon as it appears. This real-time analysis is crucial in the dynamic digital media landscape, where misinformation can spread rapidly. Rapid detection of misleading content enables timely intervention and mitigation measures.
3. *Adaptability.* NLP models can be trained and adapted to evolving linguistic patterns and to changes in the way disinformation spreads. This adaptability is essential to address the dynamic nature of language and tactics used by propagandists, ensuring that detection techniques remain effective over time.
4. *Multifaceted analysis.* NLP allows for multifaceted analysis of textual content, including sentiment analysis, semantic analysis, and context understanding. Given the different linguistic dimensions, NLP improves the ability to identify not only outright lies but also subtle forms of manipulation and biased language, often associated with propaganda.
5. *Effective pattern recognition.* NLP excels at recognizing patterns and anomalies in language structures. It can detect language inconsistencies, deviations from typical language patterns, or changes in text sentiment, providing valuable indicators of potential misinformation.
6. *Improved entity recognition.* Named Entity Recognition (NER) is a component of NLP that can identify entities such as people, organizations, and places in a text. This is especially useful for detecting inconsistencies or distortions associated with key objects, adding an extra layer of accuracy in deception detection.
7. *Automation and efficiency.* NLP automates the language analysis process, reducing the need for manual analysis. This automation increases efficiency and allows analysts to focus on more complex aspects of detecting misinformation, improving overall accuracy.

The combination of these advantages makes NLP a valuable tool in the arsenal against fakes and propaganda. Its ability to process large amounts of information, analyse content in real-time, adapt to changing linguistic nuances, and provide detailed information greatly contributes to ongoing efforts to identify and combat misinformation.

While Natural Language Processing (NLP) is a powerful tool for detecting fakes and propaganda, it is not without its limitations. Acknowledging these limitations is essential for a nuanced understanding of the challenges involved in using NLP to detect disinformation:

1. *Contextual ambiguity.* NLP models can have issues with contextual ambiguity, where the meaning of words or phrases is highly dependent on the surrounding context. Deceptive narratives often use subtle linguistic nuances that can be difficult for algorithms to accurately interpret, which can lead to potential misjudgements.

2. *Evolving nuances of language.* Language is dynamic and evolves. NLP models trained on historical data may not effectively capture new linguistic nuances and changes in expression. As disinformation tactics evolve, the inability to quickly adapt to changing linguistic trends may limit the effectiveness of NLP-based detection systems.
3. *Sarcasm and irony.* Detecting sarcasm, irony, or other forms of indirect communication poses a significant challenge to NLP models. Disinformation actors may use these linguistic tools to convey misleading narratives, and automated systems may find it difficult to accurately recognize the intended meaning without contextual cues.
4. *Lack of understanding of context.* NLP, while adept at analysing textual content, may not have a deep understanding of broader contextual factors. Understanding the intent behind a piece of information often requires an understanding of external events, cultural references, or specific contextual information that cannot be adequately captured by NLP models.
5. *Bias in training data.* NLP models are only as effective as the data they learn from. If the training data contains biases or reflects certain viewpoints, the model may inadvertently perpetuate and reinforce those biases in its assessments of the information. This poses a risk of biased results in detecting misinformation.
6. *Multilingual calls.* NLP models trained in a single language may not be equally good at recognizing misinformation in multiple languages. Adapting models to diverse language landscapes requires large training datasets and may still face difficulties in capturing the nuances of different languages.
7. *Semantic shifts.* NLP models can find it difficult to identify subtle shifts in the semantics of language, especially when disinformation actors intentionally manipulate language to convey misleading information. The ability to accurately detect semantic shifts is critical to staying ahead of new tactics used in misleading narratives.
8. *Competitive attacks.* Disinformation actors can use adversarial attacks to intentionally manipulate the results of NLP models. This can include subtle changes to the text designed to mislead the detection system, highlighting the vulnerability of NLP models to complex manipulation.

Understanding these limitations is essential for improving and empowering NLP in detecting disinformation. As this field evolves, addressing these challenges will contribute to the development of more robust and adaptive approaches to identifying misleading narratives in the evolution of the digital landscape.

4.2. Multimodal analysis

Multimodal analysis in the context of fake detection and propaganda involves the simultaneous analysis of different forms of media, such as text, images, and video. This comprehensive approach acknowledges that deceptive narratives often use a combination of textual and visual elements to convey misleading information. Multimodal analysis integrates Natural Language Processing (NLP) techniques for text analysis, image analysis for image research, and video analysis for scrutiny of video content. By assessing consistency or inconsistency between different ways, this approach improves the accuracy of detecting misleading content. Multimodal analysis also looks at metadata related to multimedia content, providing additional context for assessing authenticity. Overall, it offers a holistic view of disinformation, acknowledging the multidimensional nature of misleading narratives in contemporary media.

Multimodal analysis offers several advantages in the context of fake news detection and propaganda by taking a holistic approach that takes into account both textual and visual elements. Here are the key benefits of multimodal analysis:

1. *Textual and visual synergy.* Multimodal analysis involves the simultaneous exploration of textual and visual elements in a piece of content. By combining linguistic analysis (natural language processing or NLP) with image and video processing techniques, researchers aim to capture the synergy between textual information and accompanying visual media, recognizing that deceptive narratives often involve a coordinated interaction of words and images.
2. *Deepfake and Image Forensics.* Given the development of deepfake technology and image processing tools, multimodal analysis includes methods for identifying manipulated or synthetic content. Image forensics algorithms are used to detect changes such as deep fake faces or manipulated images, increasing the overall accuracy of deception detection.
3. *Cross-modal coherence.* Multimodal analysis seeks to assess the consistency or inconsistency between different modalities. In authentic content, textual and visual components are usually consistent. At the same time, deceptive narratives can reveal discrepancies when textual information contradicts visual elements or vice versa.
4. *Metadata checking.* Multimodal analysis goes beyond the content itself and includes the verification of metadata related to images and videos. Analysing metadata, such as timestamps, geolocation, and edit history, provides additional context for assessing the authenticity and trustworthiness of multimedia content.
5. *Recognition of patterns of different modalities.* The approach involves the development of algorithms capable of recognizing patterns of deception that span multiple modalities. This involves patterns in how certain images or visuals are combined with specific textual narratives, allowing for a more complete understanding of how propagandists create misleading messages.
6. *Social media integration.* Multimodal analysis is especially relevant in the context of social media, where misinformation is often spread through the rapid sharing of images, videos, and text content. The integration of multimodal techniques into social media analysis improves the ability to identify and counter misleading narratives on these platforms.
7. *Improved cheat indicators.* By combining signals from different modalities, multimodal analysis provides advanced indicators of deception. For example, text that claims to depict a particular event may be cross-referenced with visual content to determine whether the narrative matches the actual images, adding levels of validation in the detection process.

Multimodal analysis presents a holistic approach to detecting fakes and propaganda, taking into account the interaction of textual and visual elements. By integrating NLP techniques, image forensics, and cross-modal consistency checks, researchers aim to create more robust and comprehensive systems to identify and mitigate misleading narratives across the diverse media landscape.

While multimodal analysis is a valuable approach for detecting fakes and propaganda, it has certain limitations that affect its effectiveness. These restrictions include:

1. *Complexity of analysis.* The integration of textual, visual, and sometimes audio information increases the complexity of

the analysis. The coordination of various analytical techniques poses challenges for the development of algorithms that efficiently process and interpret information in different modalities.

2. *Resource intensity.* The analysis of multiple modalities requires significant computing resources. Multimodal analysis, especially when working with large datasets, can be resource-intensive and may require advanced computing infrastructure, limiting its availability in resource-constrained environments.
3. *New tactics of deception.* As deception tactics evolve, it can be difficult for multimodal analysis to keep up with new strategies. Adapting algorithms to detect new forms of manipulation in both textual and visual content requires ongoing research and development.
4. *Contextual challenges.* Understanding the context in which multimodal content is presented can be challenging. The distinction between deliberate deception and legitimate creative expression or satire requires a detailed understanding of the context, which can be difficult for automated systems to fully understand.
5. *The complexity of cross-modal integration.* Integrating signals from different modalities and creating meaningful connections between textual and visual elements presents challenges.

Ensuring that the analysis accurately reflects the intended message and intent of the content requires complex cross-modal integration.

6. *A rapidly evolving media landscape.* The dynamics of the media landscape, including the rapid creation and distribution of content on social media platforms, pose challenges for multimodal analysis. Real-time deception detection is becoming increasingly difficult as the volume and speed of information sharing increase.
7. *Limited generalization.* Multimodal analysis models trained on specific datasets may have trouble generalizing in different contexts or a variety of linguistic and cultural nuances. Creating robust models in different scenarios remains a pressing challenge.
8. *Competitive attacks.* Propagandists can use adversarial tactics to intentionally manipulate the output of multimodal analysis models. This poses a risk of minor changes to content aimed at avoiding detection, highlighting the vulnerability of these systems to sophisticated adversary attacks.

While multimodal analysis greatly expands the possibilities of detecting fakes and propaganda, removing these limitations is critical to improving the reliability, adaptability, and ethical considerations associated with this approach. Ongoing research and development aims to mitigate these challenges and improve the overall effectiveness of multimodal analysis in the changing disinformation landscape.

4.3. Machine learning

Machine learning is a computational approach used to analyse patterns and make predictions based on data. In the field of fake news detection and propaganda, machine learning algorithms are trained to recognize the features and characteristics associated with deceptive narratives. These algorithms study large datasets containing examples of both authentic and misleading content, allowing them to identify patterns that may indicate misinformation [4].

There are two main types of machine-learning approaches:

1. *Supervised learning.* In this approach, algorithms are trained on labelled datasets where instances of misleading content are identified. The algorithm learns to generalize based on labelled examples, making predictions based on new, unknown data.
2. *Unsupervised learning.* Unsupervised learning involves algorithms that identify patterns and anomalies in data without explicit labelling. This approach is useful for detecting misleading content without predefined examples.

Machine learning models extract functions such as linguistic structures, visual cues, or propagation patterns from data. These attributes serve as the basis for predictions or classifications regarding the authenticity of the content. Machine learning models are adaptive and able to adapt to new tactics used by those spreading misinformation.

Real-time analysis is a key advantage of machine learning in this context, allowing for rapid input evaluation to identify and counter deceptive narratives. Additionally, scalability allows machine learning models to handle the growing amount of information circulating on the internet.

Thus, machine learning for detecting fakes and propaganda involves training algorithms to recognize patterns that indicate deceptive narratives. These models adapt to new tactics, work in real-time, and scale to analyse large datasets, contributing to efforts to identify and mitigate misinformation in digital media.

Machine learning offers several advantages in the context of fake news detection and propaganda, contributing to the efficiency and effectiveness of detecting misleading narratives:

1. *Pattern recognition.* Machine learning is great at recognizing patterns in data. In the realm of fakes and propaganda, algorithms can learn patterns associated with deceptive narratives, allowing them to automatically identify similar patterns in new, invisible data.
2. *Adaptability.* Machine learning models adapt and can evolve to recognize new tactics used by disinformation spreaders. As deceptive strategies change, machine learning systems can be updated to stay ahead of evolving patterns.
3. *Real-time analysis.* Machine learning algorithms can operate in real-time, providing the ability to analyse inputs as they appear. This is crucial in a dynamic digital media environment, where timely detection of misleading content is essential for effective mitigation.
4. *Scalability*. Machine learning models can scale to process large amounts of data. Since there is a huge amount of information circulating on the internet, the scalability of machine learning approaches allows for efficient analysis of large datasets.
5. *Effectiveness.* Automated machine learning systems can quickly process and analyse data, significantly increasing the efficiency of detecting fakes and propaganda. This efficiency is essential to keep up with the rapid spread of misinformation.
6. *Ensemble methods.* Machine learning often uses ensemble techniques, combining predictions from multiple models to improve accuracy. This approach takes advantage of the strengths of different algorithms and mitigates the

weaknesses of individual models, creating more reliable detection systems.

7. *Continuous learning.* Machine learning systems can engage in continuous learning. When they encounter new data, they can adapt and update their models, ensuring that detection capabilities are effective over time.
8. *Multimodal analysis.* Machine learning facilitates multimodal analysis by processing and analysing various types of data, including text, images, and videos. This comprehensive approach increases the accuracy of identifying misleading narratives that may be used by different media formats.
9. *Evaluation metrics.* Machine learning models can be evaluated using metrics such as accuracy, memorization, and F1 score, providing a quantitative assessment of their performance. This allows for continuous improvement and optimization of detection capabilities [5].

The benefits of machine learning in detecting fakes and propaganda include its ability to recognize patterns, adapt to new tactics, work in real-time, scale efficiently, and facilitate the use of ensemble techniques. These capabilities contribute to the development of robust systems for detecting and mitigating deceptive narratives in digital media.

While machine learning is a powerful tool for detecting fakes and propaganda, it has certain limitations that affect its effectiveness in certain contexts [6][7][8][9]:

1. *Biases in training data.* Machine learning models depend heavily on the quality and representativeness of training data. If the training data contains biases or reflects certain perspectives, the model may inadvertently reinforce and reinforce these biases, leading to skewed results in detecting fakes and propaganda.
2. *Lack of contextual understanding.* Machine learning models, especially those based on narrow training data, may have trouble understanding the broader context in which the information is presented. Understanding the sarcasm, irony, or cultural nuances often used in misleading narratives can be challenging for these models.
3. *Limited generalization.* Models trained on specific datasets may have difficulty generalizing to new and diverse scenarios. The effectiveness of a machine learning model may be limited when confronted with variations in language, cultural context, or the evolution of tactics used to spread misinformation.
4. *Vulnerability to aggressive attacks.* Machine learning models can be vulnerable to aggressive attacks, where humans intentionally manipulate inputs to fool the system. Adversarial attacks in the context of fake detection and propaganda can involve subtle content changes to avoid detection.
5. *Lack of explanation.* Many machine learning models, especially complex ones such as deep neural networks, lack transparency and explanation. The black-box nature of these models can make it difficult to understand how they arrive at specific decisions, reducing trust and interpretation in the context of fake news detection and propaganda.
6. *Overtraining and undertraining.* Overtraining occurs when a model is trained too close to the training data, capturing noise rather than the underlying models. On the other hand, undertraining happens when the model is too simple to capture the relevant patterns. The balance between overtraining and undertraining is crucial for generalizing the model to new data.
7. *The dynamic nature of disinformation.* Disinformation tactics are evolving rapidly, and machine learning models can struggle to keep up with new strategies. A model trained on historical data may fail to pick up on new tactics and patterns used to spread fakes and propaganda.

8. *High resource requirements.* Some machine learning models, especially complex ones, require significant computing resources to learn and inferences. This can limit their availability in resource-constrained environments.

Understanding these limitations is essential for developing more robust and responsible machine-learning models for detecting fakes and propaganda [10][11][12]. Addressing these challenges involves continuous research, ethical considerations, and an interdisciplinary approach to improve the reliability of detection systems.

4.4. Analysis of propaganda based on emotional colouring

Sentiment analysis is a key tool in text analysis aimed at determining the emotional colouring of textual data. This method consists of determining the tone of the text as positive, negative, or neutral. Using sentiment analysis, it is possible to automatically determine how emotionally charged a certain text is, which is useful for many text analysis tasks, including the detection of propaganda materials.

Sentiment analysis is used to identify emotionally charged texts, such as propaganda materials, which are usually of a pronounced emotional nature. Propaganda texts aimed at manipulating the audience and provoking certain reactions often contain marked positive or negative sentiments that can be detected through sentiment analysis.

When analysing propaganda, it is important to take into account not only the content of the text but also its emotional load. Sentiment analysis allows you to objectively determine the emotional colouring of texts and compare them between propaganda and non-propaganda materials.

In our research, we use sentiment analysis to analyse the emotional colouring of propaganda materials compared to non-propaganda texts. We calculate the average sentiment value for each category of texts and compare their emotional characteristics.

The use of sentiment analysis is an important step in identifying propaganda materials and understanding their emotional impact on the audience. This method helps to make an objective assessment of texts and identify their potential impact on the perception of information.

```python
def main():
    (propaganda, non_propaganda) = get_datasets()

    (propaganda_emotions, propaganda_positive, propaganda_negative, propaganda_subjective, propaganda_mixes) = analyze_emotion(propaganda)
    (non_propaganda_emotions, non_propaganda_positive, non_propaganda_negative, non_propaganda_subjective, non_propaganda_mixes) = analyze_emotion(non_propaganda)

    propaganda_result_emotions = sum(propaganda_emotions) / len(propaganda_emotions)
    non_propaganda_result_emotions = sum(non_propaganda_emotions) / len(non_propaganda_emotions)

    propaganda_result_positive = sum(propaganda_positive) / len(propaganda_positive)
    non_propaganda_result_positive = sum(non_propaganda_positive) / len(non_propaganda_positive)

    propaganda_result_negative = sum(propaganda_negative) / len(propaganda_negative)
    non_propaganda_result_negative = sum(non_propaganda_negative) / len(non_propaganda_negative)

    propaganda_result_subjective = sum(propaganda_subjective) / len(propaganda_subjective)
    non_propaganda_result_subjective = sum(non_propaganda_subjective) / len(non_propaganda_subjective)

    propaganda_result_mix = sum(propaganda_mixes) / len(propaganda_mixes)
    non_propaganda_result_mix = sum(non_propaganda_mixes) / len(non_propaganda_mixes)
```

Fig. 1. Program for analysing the emotional colouring of propaganda and non-propaganda materials

A program has been developed to analyse texts from two different datasets (Fig. 1): propaganda news and non-propaganda news. This program uses the TextBlob library to analyse the emotional colouring of each text and obtain an objective assessment of the emotional tonic.

The program analyses each text to determine its emotional colouring, positive and negative emotional colouring, the degree of subjectivity and complex emotional colouring. To do this, she uses the methods of the TextBlob library, which traces the emotional tone of the text and determines its characteristics.

- *Emotion* (Fig. 2): The average emotional value for propaganda news is 0.151 and for non-propaganda news is 0.116.
- *Subjectivity* (Fig. 3): The average degree of subjectivity for propaganda news is 0.365 and for non-propaganda news is 0.283.
- *Positive emotion* (Fig. 4): The average value of positive emotion for propaganda news is 0.087 and for non-propaganda news is 0.082.
- *Negative emotion* (Fig. 5): The average negative emotion for propaganda news is 0.064 and for non-propaganda news is 0.034.
- *Complex* (Fig. 6): The average value of the complex emotional colouring for propaganda news is 0.021, and for non-propaganda news - 0.010.

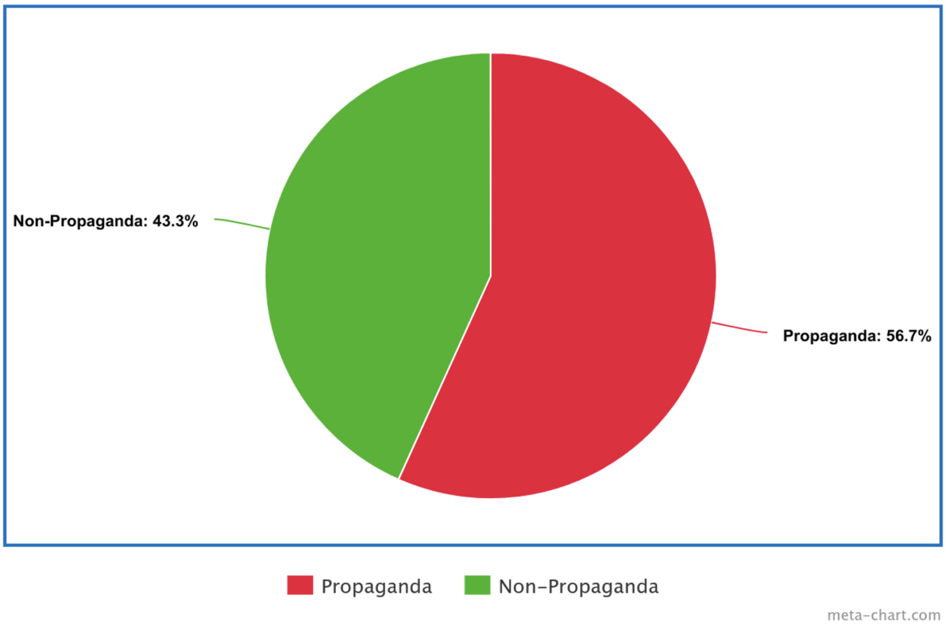

Fig. 2. Emotionality

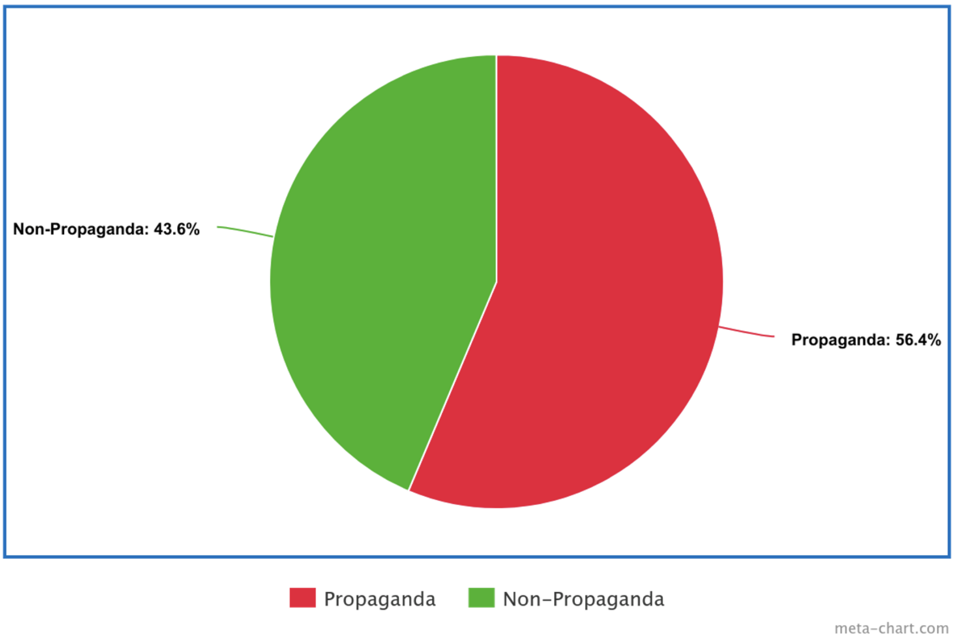

Fig. 3. Subjectivity

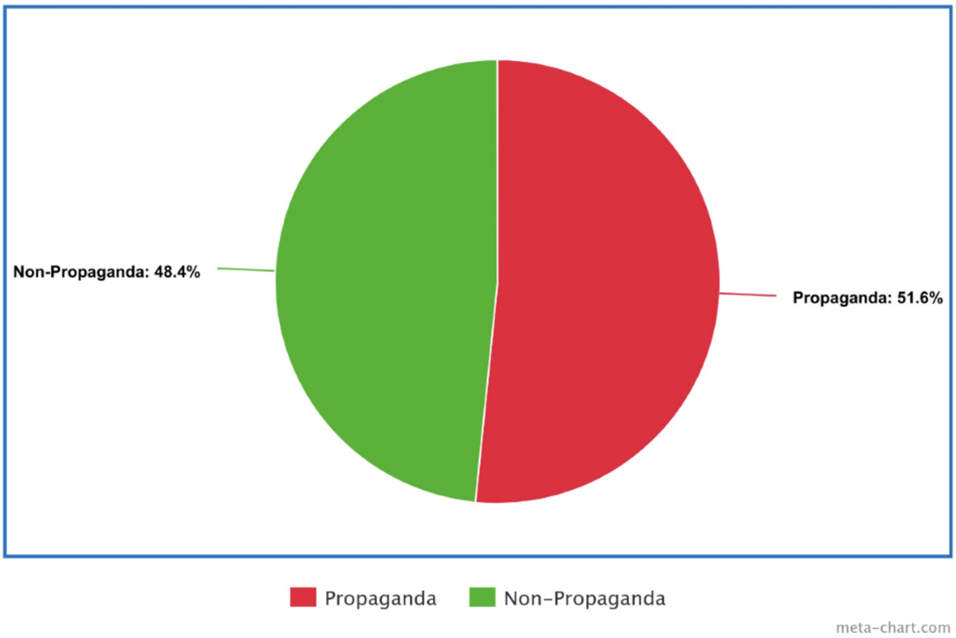

Fig. 4. Positive emotions

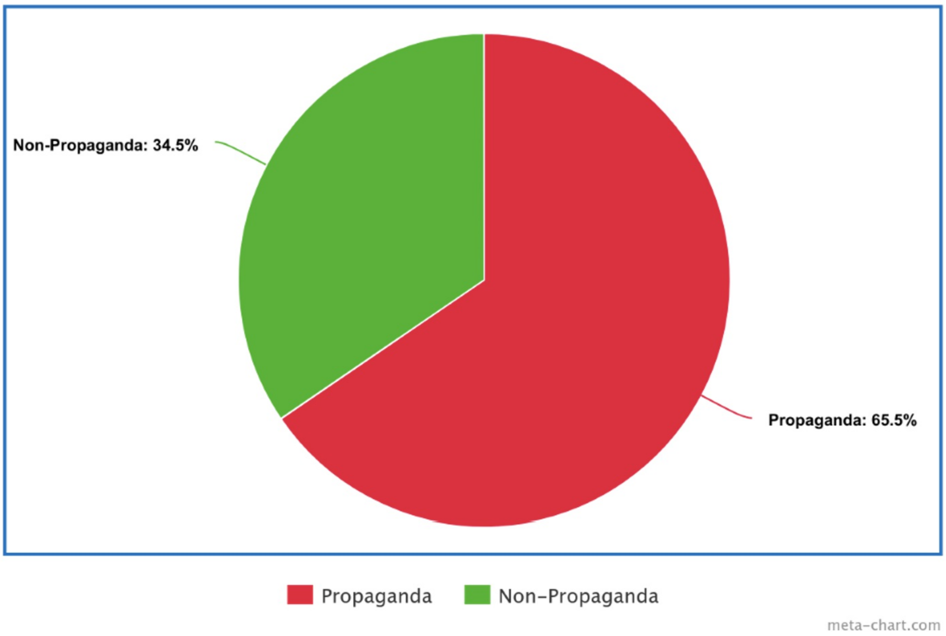

Fig. 5. Negative emotions

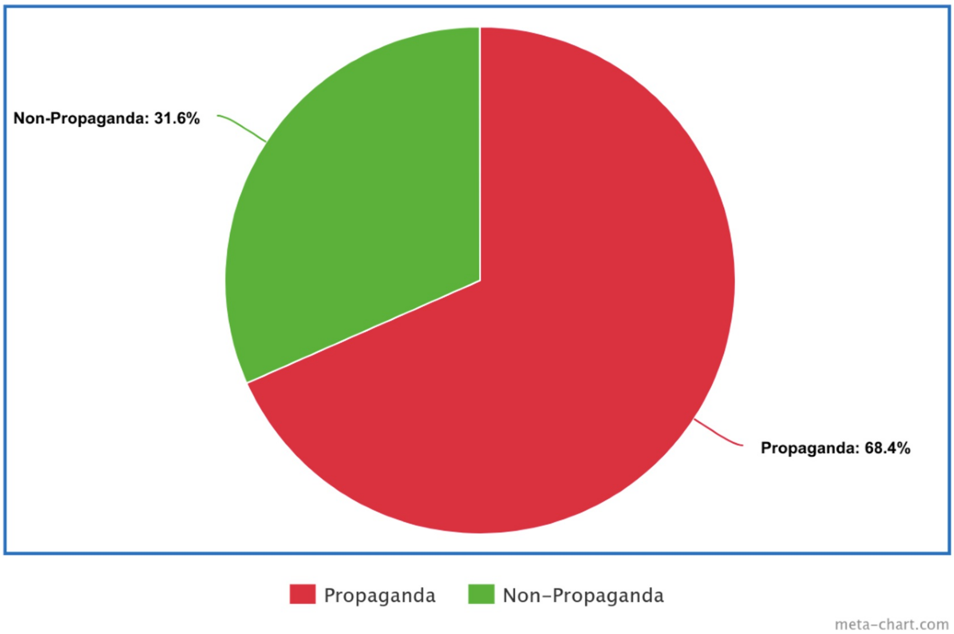

Fig. 6. Comprehensive assessment

The results of the analysis of the emotional colouring of the texts indicate several key differences between propaganda and non-propaganda materials. In particular, there is much more negative emotional connotation in propaganda news compared to non-propaganda news. The average value of negative emotional connotation for propaganda news is 0.064, while for non-propaganda news it is only 0.034. This indicates that propaganda materials tend to accentuate negative aspects, which may be aimed at forming a certain negative opinion or emotional reaction in readers.

However, as for the positive emotional colouring, it is observed that its level in propaganda and non-propaganda news is approximately the same. The average value of positive emotional colouring for propaganda news is 0.087 and for non-propaganda news - 0.082. This may indicate that positive emotions are being used in propaganda materials to increase attractiveness and attract the attention of the audience.

The biggest difference is observed in the indicator of complex emotional colouring. In propaganda news, this figure is twice as high as in non-propaganda materials. This may indicate that the emotional colouring of propaganda texts is more complex and includes more different aspects, such as negativity, emotionality and subjectivity. This difference can be used to develop more accurate algorithms for detecting and analyzing propaganda materials in texts, which will improve the efficiency of identifying such materials and understanding their impact on the audience.

5. Conclusions

In the digital age, the pervasive impact of fakes and propaganda requires a sophisticated and multifaceted approach to detect and mitigate. This article discusses various techniques, with a particular focus on Natural Language Processing (NLP), multimodal analysis, and machine learning in the context of detecting deceptive narratives. Each method provides unique strengths for achieving the main goal of protecting the integrity of information dissemination.

Natural language processing research has highlighted its effectiveness in analysing language patterns, while multimodal analysis recognizes the multimedia nature of modern disinformation. Machine learning, with its adaptability and real-time capabilities, provides a dynamic tool for recognizing evolving tactics.

Interdisciplinary collaboration is of paramount importance. Bringing together technologists, sociologists, policymakers, and media professionals, a more comprehensive understanding of the socio-political and ethical dimensions of disinformation is emerging.

Contextual issues related to language, cultural nuances, and the rapidly evolving landscape of disinformation highlight the need for ongoing research to improve contextual understanding of detection systems.

Given the dynamic nature of disinformation tactics, ongoing research is needed to develop models that can dynamically adapt to new strategies. This involves creating systems that stay ahead of evolving patterns through continuous training and updates.

Improving the integration of different ways of detecting disinformation is crucial. Research efforts should explore techniques that seamlessly combine textual, visual, and possibly audio information to create more coherent detection systems.

The emotional connotation in propaganda materials is more pronounced in non-propaganda materials. Propaganda news tends to make more negative emotional statements, which can be used to manipulate audiences. There is also a greater complexity of emotional colouring in propaganda materials.

In conclusion, the fight against fakes and propaganda requires a comprehensive and evolving strategy. The synergy of technological advances, ethical considerations, interdisciplinary collaboration, and user empowerment is essential. As we move forward, it is important to view the detection of fakes and propaganda not as a standalone challenge, but as a dynamic area where ongoing research and development are key to staying ahead of those who seek to manipulate information for their ends. Only through concerted and adaptive efforts can we hope to build a resilient defence against the pervasive influence of misleading narratives in our digital age.

References

1. ^*Haidaichuk, V., and Smal, M. (2021). "Metodichni pidkhody do vyznachennia fakekiv ta propaganda v neti Internet" [Methodological approaches to the detection of fakes and propaganda on the Internet]. Scientific Bulletin of Uzhhorod National University. Series: Philology. Issue. 2, pp. 27-31.*
2. ^*Zhang Z., Chen Y., Chen L. (2020). Methods for detecting fake news: A survey. ACM Computing Surveys. Vol. 53. No. 6. P. 1-37.*
3. ^*Yakovleva, O. (2021). Methods of detecting fakes and propaganda in text data. Scientific Bulletin of the Dnipro University. Olesya Gonchara. Series: Computer Science. Issue. 2, pp. 28-32.*
4. ^*Gupta M., Mishra A., Kumar V. (2020). A survey on fake news detection using machine learning techniques. Journal*

of Big Data. Vol. 7. No. 1. P. 1-21.

5. ^*Methods of detecting fakes and propaganda in artificial intelligence. (2020). Scientific Bulletin of Lviv Polytechnic National University. Series: Informatics. Issue. 2, pp. 113-117.*
6. ^*Oliinyk, V. A., Vysotska, V., Burov, Y., Mykich, K., & Fernandes, V. B. (2020). Propaganda Detection in Text Data Based on NLP and Machine Learning, volume Vol-2631 of CEUR workshop proceedings, 2020, pp. 132-144.*
7. ^*Raturi, S., Mishra, A. K., & Maji, S. (2022). Fake News Detection Using Machine Learning. In DeepFakes (pp. 121-133). CRC Press.*
8. ^*Mykytiuk, A., Vysotska, V., Markiv, O., Chyrun, L., & Pelekh, Y. (2023). Technology of Fake News Recognition Based on Machine Learning Methods. CEUR Workshop Proceedings, Vol-3387, 311-330.*
9. ^*Vysotska, V., Mazepa, S., Chyrun, L., Brodyak, O., Shakleina, I., & Schuchmann, V. (2022, November). NLP tool for extracting relevant information from criminal reports or fakes/propaganda content. In 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT) (pp. 93-98). IEEE. doi: 10.1109/CSIT56902.2022.10000563.*
10. ^*Dar, R. A., & Hashmy, R. (2023). A Survey on COVID-19 related Fake News Detection using Machine Learning Models, CEUR Workshop Proceedings, Vol-3426, 36-46.*
11. ^*Afanasieva, I., Golian, N., Golian, V., Khovrat, A., & Onyshchenko, K. (2023) Application of Neural Networks to Identify of Fake News, CEUR Workshop Proceedings, Vol-3396, 346-358.*
12. ^*hupta, A., Barmak, O., Wierzbicki, A., & Skrypnyk, T. (2023). An Adaptive Approach to Detecting Fake News Based on Generalized Text Features, CEUR Workshop Proceedings, Vol-3387, 300-310.*