LIMITATION OF BLACK-MARKET MEDIA INFLUENCE THROUGH DESIGNING LLM TOOL FOR CONTENT ANALYSIS

Boyana Ivanova, Rumyana Bratanova*

University of Ruse "Angel Kanchev", Ruse
Bulgaria

* Corresponding Author, e-mail: bivanova@uni-ruse.bg

Abstract: The meaning and importance of restricting black-market media are identified. A clear definition of Large Language Models (LLMs) as part of Artificial Intelligence (AI) is given. These two terms are associated with preventing the negative influence of media on the masses. The research blends two trending topics - Generative AI and content analysis through the prism of Artificial Intelligence. A classification of software tools that use Generative AI in the context of content analysis is given. The paper proposes a methodology for designing and developing an LLM tool. All the included steps are described in detail. The key steps and approaches are described as constructive parts of the tool. A list of best-practice examples of applications that use content analysis with Large Language Models is presented. As a result, a combination of innovative technologies, modern approaches, and defined challenges in solving problems with interdisciplinary character is presented.

Key words: large language models, artificial intelligence, content analysis, interdisciplinary, black market.

1. INTRODUCTION

The term "fake news" has become increasingly popular in recent years. While this term has numerous synonyms - disinformation campaigns, cyber propaganda, cognitive hacking and information warfare - it represents only one facet of a more substantial problem: the manipulation of public opinion to influence real outcomes. Thanks to connectivity and digital platforms that facilitate the spread of information, there are no longer traditional limitations such as geographical boundaries and restrictions of time and distance.

In the public and private sectors, a substantial volume of information is collected and processed, including the personal data of employees, customers, collaborators, and other stakeholders. In the context of the digital society, a significant proportion of this information is collected, processed, and stored in electronic form, which necessitates

specific requirements for ensuring computer and network security as integral components of the broader concept of information security [1].

The question of how technologies could help prevent black-market media influence has a simple answer - Large Language Models. LLMs are essentially based on machine learning. Machine learning is a branch of artificial intelligence and refers to the practice of feeding a program with large amounts of data to train the program to recognize features of this data without human intervention. The utilization of Large Language Models to mitigate the proliferation of black-market media represents an innovative approach encompassing the classification of LLM applications and existing tools, the design of LLM tools with a comprehensive, elucidated methodology, the delineation of primary considerations on the implementation of such software and identifying benefits, challenges, and future trends in preventing from black-market media influence with AI integration.

2. RELATED WORK

The business model is a crucial part of business architecture, guiding the development of software that aligns with the business's specifics and functionalities [2]. The Internet of Things, Industrial Internet of Things, Big Data analytics, and Machine Learning technologies have improved prediction and control of various characteristics. Industry 4.0 technologies, such as IoT, Big Data analytics, and Machine Learning, are expected to enhance productivity, reduce costs, and reduce customer service time. Predictive maintenance in industrial enterprises contributes to increased productivity and reduced downtime. Technologies like Big Data analytics, Machine Learning, and deep learning enable the creation of models capable of generating accurate enterprise operations [3]. LLMs use deep learning, which can recognize distinctions without human intervention, but require training and fine-tuning for specific tasks.

2.1. LLMs in the context of black-market media prevention

Using LLMs to prevent the proliferation of black-market media can be an innovative approach by using some techniques shown in Figure 1 and the following literature review is based on these techniques and methods.

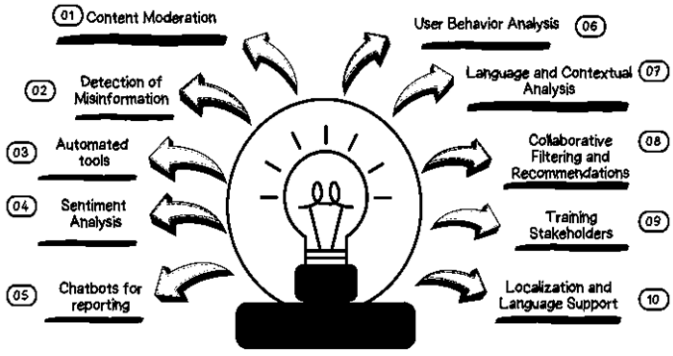

Figure 1. Black-market media prevention by using LLMs

Content moderation: LLMs can be trained to identify and filter out harmful or illegal content on media platforms. The models can remove content that violates community guidelines or legal norms by analyzing text, images, and videos. The natural language understanding capabilities of LLMs open opportunities to use LLMs for online moderation [4].

Detecting misinformation: LLMs can help identify misinformation that may be linked to black market activity by analyzing claims in online articles or social media posts and cross-referencing them with verified databases and sources. Automated reporting tools: software tools that allow users to report suspicious content easily can be developed with LLMs. These tools can categorize reports based on keywords or patterns associated with black market activity.

Sentiment Analysis: LLMs can facilitate the monitoring of public sentiment, and the identification of emerging trends related to black market activities. This information can be utilized to allocate law enforcement resources more efficiently. Chatbots for Education and Reporting: The deployment of LLM-powered chatbots on websites can serve to educate users about the risks associated with black market media and methods to avoid it. Research conducted among students in higher education showed that LLMs Stable diffusion approach are good to generating images for different topics and better understanding [5].

Additionally, these chatbots can guide on reporting illegal activities or suspicious content. User Behaviour Analysis: LLMs can analyze user interactions on platforms to identify behavioural patterns that may indicate engagement with black market media. For instance, they can identify users who frequently share or engage with flagged content. Language and Contextual Analysis: LLMs can determine when content is inappropriately marketed or misused for illicit purposes including the identification of coded language often used in black markets. An example of contextual analysis could be smishing detection. A paper presents a smishing (SMS phishing) detection model using NLP and ML techniques. The developed model uses NLP algorithms to analyze text-based messages, examining linguistic patterns and contextual cues that indicate smishing attempts. With the help of ML algorithms, the model learns to distinguish between legitimate (non-smishing) and fraudulent messages (smishing) and dynamically adapts to evolving smishing tactics.

Despite challenges such as limited access to data and lack of awareness among respondents, the study makes an important contribution to the state of knowledge by identifying linguistic markers that clearly indicate smishing messages, evaluating the model's performance, and highlighting the robust performance of the Random Forest model in accurately identifying smishing attempts [6]. Classifying unstructured text data written in natural languages is a complex task, particularly in cases involving extensive datasets with multiple languages. In a study, the researcher investigated the application of Long Short-Term Neural Network (LSTM) in developing a classification model capable of learning text patterns and categorizing English and Tagalog-based complaints, feedback, and commendations from customers in the context of a state university in the Philippines. The results demonstrated that the LSTM achieved its optimal training accuracy of 91.67% and elapsed time of 34s when configured with 50 word embedding size and 50 hidden units. The investigation revealed that a lower

number of hidden units in the network correlates with higher classification accuracy and faster training time, while word embedding size exhibits no correlation to the classification performance. The study determined that adjusting the word embedding size has no significant correlation with training accuracy and elapsed time. The proposed text classification model demonstrated effectiveness with 95% classification accuracy during testing implementation [7].

Collaborative Filtering and Recommendations: Through the analysis of user behaviour and context, LLMs can be implemented in recommendations that discourage engagement with black market content. Training Stakeholders: LLMs can assist in the creation of training materials and resources for law enforcement agencies, social media companies, and educators on how to recognize and beat the black-market media effectively. Localization and Language Support: In the context of the black market, LLMs can be adapted to improve the detection of activities in different regions, thus facilitating more targeted prevention efforts.

While LLMs can serve as powerful tools in preventing black market activities in media, it is imperative to utilize them responsibly and ethically, considering privacy concerns and the potential for bias in AI systems. The implementation of these solutions should also involve collaboration among technology developers, policymakers, and communities.

2.2. LLMs content analysis tools classification

The research aims to design a tool using Large Language Models (LLMs) to prevent mass black market media. The paper presents a classification of existing content analysis tools using Large Language Models and proposes a detailed methodology for designing a prototype of a such software. Content analysis is one of the most commonly used techniques for finding black market media.

LLMs are a transformative approach to understanding and categorizing texts, using deep learning to process large amounts of data and generate human-like answers. They are widely used across industries for analyzing and interpreting unstructured data, including textual, social media, and multimedia content. The study presents a At Figure 2 is shown a classification of existing content analysis tools using AI techniques, identifying seven categories that combine business and IT to create the synergy necessary for deep and qualitative content analysis. The categories are formed depending on the application in content analysis.

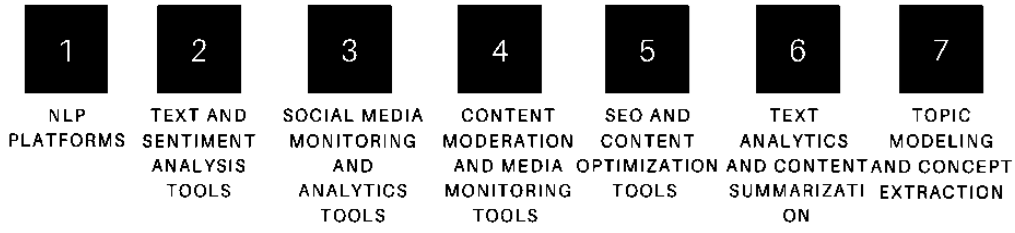

Figure 2. AI Content Analysis Tools categories

The first category comprises NLP platforms like Google Cloud Natural Language API, Microsoft Azure Text Analytics, and IBM Watson Natural Language Understanding. The second focuses on text and sentiment analysis tools, featuring applications such as MonkeyLearn, Lexalytics, and RapidMiner. The third category highlights social media monitoring tools, including Brandwatch, Hootsuite Insights, and Sprinklr. Content moderation and media monitoring are represented by Clarifai, PimEyes, and Meltwater in the fourth category. A proposed framework for content moderation stresses the importance of accuracy, efficiency, and stakeholder engagement [8]. The fifth category pertains to SEO and content optimization tools like MarketMuse, Frase.io, and Surfer SEO. Text analytics tools, including QuillBot and Grammarly, fall into the sixth category. An interesting study has explored spam email detection using machine learning and NLP, demonstrating advanced accuracy via transformer-based models like BERT [9]. The seventh category includes topic modelling tools such as Gensim and Spacy for analyzing diverse content.

3. PROBLEM DESCRIPTION AND PROPOSED SOLUTION

The presented classification of artificial intelligence tools for content analysis that focuses on protecting the public from the influence of black media that could be integrated into a tool specifically designed to address the needs of protecting the public from the influence of black media, which constitutes the focus of this study. A methodology is proposed that ensures high-quality development, incorporation of innovative technologies, integration with external systems, adherence to programming conventions, and optimization of software distribution and content performance. The methodology includes twelve steps for the design and development of a content analysis tool using Large Language Models. Steps 7 to 9 are iterative, i.e. they are repeated until the actual result matches the expected result (Figure 3).

Step 1. A Large Language Model (LLM) content analysis tool can help prevent black-market media by analyzing text-based content. It includes functions like text classification, sentiment analysis, entity recognition, keyword extraction, topic modelling, and contextual understanding. Text classification organizes documents into predefined categories, identifying specific topics. Sentiment analysis determines the sentiment of a text, allowing for trend analysis. Entity recognition identifies people, organizations, places, and dates in the text, extracting relevant information. Keyword extraction highlights important keywords for indexing or searching. Topic modelling groups documents by topic for improved organization and searchability. LLMs can distinguish between literal and figurative language, identify plagiarism, and analyze user behaviour. Notification systems provide notifications for specific keywords, trends, or sentiment changes. Reporting and dashboards provide dashboards for monitoring content performance metrics. These functionalities enhance the capacity to analyze, comprehend, and manage content across various domains, contributing to informed decision-making and strategic planning.

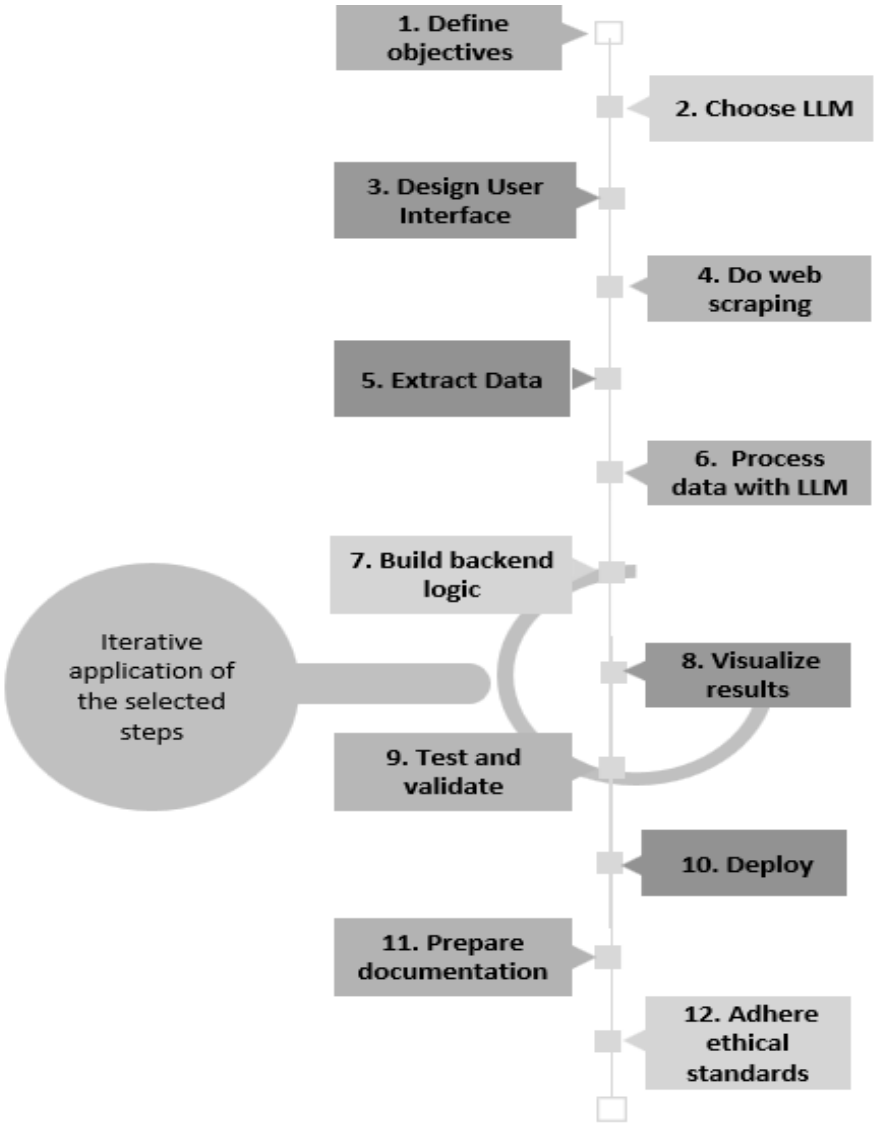

Figure 3. Methodology of designing LLM content analysis tool

Step 2. Selecting the right LLM for content analysis depends on factors like model size, fine-tuning capabilities, and availability of Application Programming Interfaces (APIs). Several software solutions, libraries, and platforms use LLMs and offer various functionalities for content analysis. Examples include OpenAI (ChatGPT, API), Google Cloud Natural Language API, IBM Watson Natural Language Understanding software, Microsoft Azure Text Analytics, Hugging Face Transformers, Amazon Comprehend, TextRazor, MonkeyLearn, Lexalytics, Rasa, ContentSquare, SAS Text Analytic, and Aylien. These platforms offer a combination of functionalities suited for content analysis

and can be adapted or integrated depending on specific use cases and tool requirements. Examples include OpenAI's ChatGPT, Google Cloud's Natural Language API, IBM Watson's Natural Language Understanding, Microsoft Azure Text Analytics, Hugging Face Transformers, Amazon Comprehend, TextRazor, MonkeyLearn, Lexalytics, Rasa, ContentSquare, SAS Text Analytic, and Aylien. Each platform or tool offers a combination of functionalities suited to content analysis and can be adapted or integrated depending on specific use cases and tool requirements.

Step 3. A user-friendly interface is essential for a positive user experience. To achieve this, conduct user research, create user personas, ensure simple and intuitive navigation, provide a responsive design with flexible layouts, images and CSS media queries, and use high-contrast colours to highlight key elements. Prioritize essential features and information and consider a design system with standardized styles and guidelines for developers. Provide feedback in the form of messages, animations or error messages when users perform actions and offer actionable steps for resolution. Ensure the user interface is accessible to people with disabilities by providing colour contrast, keyboard navigation, screen reader compatibility and meaningful labels. Optimize performance by optimizing images, scripts and resources for fast load times. Keep the user focused by avoiding unnecessary pop-ups and modal windows and providing clear notifications when the user completes multi-step processes. By following these best practices, a user-friendly interface can enhance satisfaction, reduce frustration, and encourage continued engagement with the application or website.

Step 4. Web scraping modules like BeautifulSoup or Scrapy can be used to extract content from websites. These modules can also help understand the structure of websites and generate scripts for data extraction. The workflow involves examining HTML code to identify the desired data location, using free tools to separate elements, and identifying relevant tags and attributes. For dynamic content, LLMs can suggest using Selenium software to interact with the browser. If a website uses pagination, a script can be developed to create a loop to navigate through pages and extract data. After data extraction, information can be synthesized, calculations performed, or analyses generated. It is important to consult the robots.txt file of the website and consider the ethical and legal implications of web scraping.

Step 5. Data processing with LLM consists of two main activities: pre-processing, which involves cleaning and preparing the extracted text (removing HTML tags and special characters), and analysis functions, which may include sentiment analysis (utilizing LLM to determine the sentiment of the text), keyword extraction (implementing a function to identify the most relevant keywords using LLM), and topic modeling (employing LLM to identify topics or themes within the content). This marks the beginning of the iterative phase of the methodology, where the steps are repeated until results meeting the acceptance criteria of the test strategy are achieved. Three approaches can be applied for sentiment analysis: TextBlob and spaCy libraries, VADER (Valence Aware Dictionary and sEntiment Reasoner), and transformer models like BERT or DistilBERT. These models are trained on extensive datasets and can capture more nuanced sentiments.

Step 6. The LLM must be integrated with backend logic for a server-side application, which can be done using various frameworks and languages. Python (Flask)

is a popular choice for handling requests, processing data, and returning results. To develop a simple backend logic, install Flask, create a Python script, and set up a simple Flask app. The app will respond to incoming requests, process data, and return responses. For more complex processing, the logic can be extended within route functions. Flask allows for JSON format, making it easy to return results. The API can be tested using tools like Postman or Curl. Once working locally, the server-side logic can be deployed to a cloud service or connected to a front-end application. Basic backend logic can be integrated with databases using libraries like SQLAlchemy or PyMongo, authentication & authorization modules with Flask-JWT-Extended, and file uploads and processing with Flask's request. files function.

Step 7. Several libraries are available for visualizing results from tasks such as sentiment analysis, text generation, named entity recognition (NER), topic modelling, and related applications. These libraries facilitate the creation of graphs, charts, and more complex visualizations to derive insights from LLM-generated data (Table 1).

Table 1. - Libraries used for visualizing data

Library	***Usage***	***Features***
Matplotlib	General-purpose plotting – bar charts, line plots, histograms	Widely used; highly customizable, integration with many other libraries
Seaborn	Statistical data visualization, heatmaps, histograms, pair plots	Build on top of Matplotlib, providing more aesthetic plots with less code
Plotly	Interactive plots – 3D plots, maps, dashboards	Highly interactive, could be embedded in web applications
WordCloud	Visualization of word frequency in text	Showcasing dominant words in LLM-generated text or topic modelling results
Altair	Declarative statistical visualization – scatter plots, histograms, bar charts	Easy to use; good for quick interactive visualization
SpaCy's Displacy	Visualization of NLP tasks – NER, dependency parsing	Integrated with spaCy; could output HTML visualization
UMAP	Dimensionality reduction and visualization of embeddings	Visualization of high-dimensional data in 2D or 3D space interactive
Bokeh	Interactive visualizations	Rich interactive plots

Step 8. The testing process is crucial LLMs to ensure their functionality and accuracy. It involves validating results against benchmarks or data sets, ensuring they meet deployment criteria. Key aspects include accuracy and coherence. LLMs use five types of software tests: unit testing, functional testing, regression testing, performance testing, and security testing. Unit tests test the smallest testable parts of an application, while functional tests evaluate multiple unit tests for a specific use case. Regression

testing evaluates the LLM with the same set of test cases to prevent erroneous changes. Performance testing focuses on the LLM's efficiency, while security testing ensures the robustness and security of LLM-based applications. Compliance with legal, regulatory, ethical, privacy, and fair handling of harmful biases is essential for the integrity of LLM-based applications.

Step 9. LLM tool deployment involves various components and procedures to ensure efficient operation. There are three methodologies: Cloud Deployment, Managed Services for LLMs, and On-Premises/Local Deployment. Cloud deployment is suitable for organizations with data control or cloud expenses. On-premises deployment is suitable for those needing complete control. Local deployment requires sufficient GPU/CPU resources for model inference. Containerization tools like Docker and Kubernetes can facilitate consistent deployment on local machines.

Step 10. A comprehensive user manual is essential for the development and deployment of an LLM tool. It includes an introduction, getting started guide, API documentation, user interface guide, use case examples, performance and resource considerations, security considerations, error handling, versioning and updates, licensing and terms of use, and support channels. This helps users maximize the tool's benefits, solve problems, and work effectively with the support team.

Step 11. Ethical standards for content analysis using LLMs are essential to ensure responsible and fair application. These standards focus on transparency, fairness, privacy, accountability, and harm minimization. Key ethical considerations include explainability, fairness, privacy, accountability, intellectual property, and transparency in data use. These standards align with broader societal values like fairness, responsibility, and harm reduction, ensuring that LLMs used for content analysis are consistent with these values.

4. DISCUSSION AND FUTURE WORK

Implementing innovative technologies, such as Large Language Models, has its benefits, challenges, and future in many fields. Some of the most outstanding benefits of using LLMs for content analysis are scalability, contextual understanding, customizability and automation. Scalability feature could process huge amounts of data in real-time, making them suitable for large-scale content analysis tasks. The context, subtle meanings, implicit messages, and nuances in conversations could be deeply understand by LLMs. Customizability shows the ability of a lot of industries to adopt the technology by customizing the types of content (e.g. legal, financial, medical) using fine-tuning and improving accuracy and relevance in these areas. LLMs enable the automation of repetitive tasks such as labelling, summarizing, and categorizing data, reducing manual work and enabling more efficient workflows.

The identified challenges and considerations in using LLMs for content analysis refer to bias and fairness where LLMs can amplify the biases that are presented in their training data. Insurance of fairness and mitigating bias in content analysis tasks are critical; interpretability: LLMs can be used as black-box models that means it could be difficult to understand how they made certain conclusions, the lack of transparency is a challenge in highly regulated industries; data privacy when analyzing sensitive data such

as customer feedback or medical records, the compliance with data privacy regulations (like GDPR) is critical; training costs: Fine-tuning and training LLMs on domain-specific data can be resource-intensive, although pre-trained models like GPT-4 and BERT have significantly reduced barriers to entry.

Future trends in using LLMs for content analysis regard the usage in different fields and industries, the attitude of people that use the LLMs. The LLMs could be utilized to develop interactive tools that enable users to search data in natural language by providing an intuitive interface for content analysis. With advancements in computational efficiency, LLMs may facilitate real-time analysis of content, particularly on social media platforms or news aggregators. LLMs are increasingly being integrated with vision and speech models, leading to more comprehensive analysis that incorporates text, video, and images. The technology represents a powerful tool in content analysis, significantly expanding the capacity to derive insights from vast amounts of text and complex datasets.

5. CONCLUSION

LLMs will inevitably become even more used in all aspects of life. This paper aims to recognize the stages of implementing LLMs in the field of content analysis.

The implementation of large language models in content analysis offers significant advancements in both the efficiency and depth of data interpretation.

Large Language Models can do tasks such as content categorization, sentiment analysis, and topic modelling with unprecedented accuracy. The reason is their capacity to process large amounts of unstructured data. This ability to understand context and complex language patterns enhances qualitative and quantitative research approaches, facilitating a comprehensive understanding of textual data. Integrating LLMs can reduce the manual work required for content analysis by enabling more rapid data processing and scalable analyses.

This research paper proposes a methodology that effectively implements LLMs in content analysis to mitigate the dissemination of black-market media. Designing and developing a large language model tool in the context of content analysis provides a robust instrument for understanding and preventing the spread of illicit information and illegal content because LLMs can efficiently process and analyze vast amounts of unregulated or underground media, enabling the identification of trends, key actors, and networks involved in black market activities. Furthermore, LLMs can assist in detecting coded language and other obfuscation techniques frequently employed in black market media, which may be challenging for human analysts to interpret. This capability enhances law enforcement and policy efforts to monitor and disrupt illicit activities. However, significant ethical challenges concern privacy, surveillance, and potential misuse of these technologies in monitoring online mass behaviour.

In conclusion, while LLMs have the potential to analyze content in black market media, their utilization must be governed by strict ethical guidelines to ensure responsible implementation. Future research will focus on enhancing the accuracy and fairness of these models in analyzing illicit content and developing procedures to protect individual privacy and prevent misuse. This paper will serve as a foundation for

developing a software tool using the identified stages and steps and will follow the successful implementation of the proposed methodology, the identified trends, and challenges.

ACKNOWLEDGEMENT

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.013-0001-C01.

REFERENCES

[1] Noninska I., Romansky R. Features of e-business and e-commerce with a reflection on the privacy in the digital age. *International Journal on Information Technologies and Security*, vol.15, no.3, 2023, pp. 115-126. https://doi.org/10.59035/VOTY8866.

[2] Shoilekova, K. (2023). Business model as a basic building block of business architecture. Software engineering application in systems design. CoMeSySo *2022. Lecture Notes in Networks and Systems*, vol 596. Springer, Cham. pp. 563–567. https://doi.org/10.1007/978-3-031-21435-6_48.

[3] Gritsyuk K. M., Gritsyuk V.I. Technologies and tools used for prediction of characteristics of activity of industrial enterprises. *International Journal on Information Technologies and Security*, vol.12, no.4, 2020, pp. 47-63.

[4] Kolla M., Salunkhe S., Chandrasekharan E., Saha K. LLM-Mod: Can Large Language Models Assist Content Moderation. *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24).* Association for Computing Machinery, New York, NY, USA, Article 217, pp 1–8. https://doi.org/10.1145/3613905.3650828.

[5] Kozov V., Ivanova G. Atanasova D., Practical application of AI and Large Language Models in software engineering education, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.15, issue 1, 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150168.

[6] Aaron Zimba A., Phiri K. O., Kashale C., Phiri M.N. A machine learning and natural language processing-based smishing detection model for mobile money transactions. *International Journal on Information Technologies and Security*, vol.16, no.3, 2024, pp. 69-80. https://doi.org/10.59035/UEEN6450.

[7] Corpuz R.S.A. An application method of long short-term memory neural network in classifying English and Tagalog-based customer complaints, feedback, and commendations. *International Journal on Information Technologies and Security*, vol.13, no.1, 2021, pp. 89-101.

[8] Huang T. *Content Moderation by LLM: From Accuracy to Legitimacy*. 2024. https://doi.org/10.48550/arXiv.2409.03219

[9] Alshawi B., Munsh A., Alotaibi M., Alturki R., Allheeib N. Classification of SPAM mail utilizing machine learning and deep learning techniques. *International Journal on. Information Technologies and Security*, vol.16, no.2, 2024, pp. 71-82. https://doi.org/10.59035/FPKO7430

Information about the authors:

Boyana Ivanova - an Assistant professor in the Department of Informatics and Information Technologies at the University of Ruse and PhD in Computer Science. Her scientific and research interests are in the area of Artificial Intelligence, Mixed Reality, Cybersecurity and Education Technologies.

Rumyana Bratanova, PhD - a researcher at University of Ruse in the field of National security and a Deputy director of the educational department of University hospital MEDIKA, Ruse, Bulgaria. Her scientific and research interests are in the area of Artificial Intelligence, Robotics, Healthcare.

Manuscript received on 16 November 2024