REVIEW
of the official reviewer for the dissertation work of
Karyukin Vladislav Igorevich on the topic "The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan", granted for the degree of Doctor of Philosophy (PhD) in the specialty "6D070300 – Information systems".

№ p/p	Criteria	Compliance with the criteria (you must mark one of the answer options a)	Substantiation of the position of the official reviewer
1.	The dissertation's topic (as of its approval date) corresponds to the directions of the development of science and/or state programs.	1.1 Compliance with priority areas of science development or state programs: the topic of the dissertation work, "The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan" corresponds to priority areas of science, such as intelligent systems, artificial intelligence, social research, and information technologies.	Artificial intelligence methods, especially natural language processing (NLP), machine learning, and deep learning models, is a key area in scientific research. These methods can significantly improve the system's ability to analyze and understand data from the media space. The analysis of social sentiment also requires the application of social sciences to understand and evaluate society's opinions, moods, and reactions to various socially oriented programs and policies of the state. Developing an integrated system that includes the collection, processing, and analysis of Big Data is also very important. The development of such a system will effectively manage and analyze huge amounts of data, which opens up new opportunities for research and understanding of the social mood of society in the media space.
		1) The dissertation was completed within the framework of a project or target program financed from the state budget (indicate the name and number of the project or program) 2) The dissertation was completed within the framework of another state program (indicate the name of the program) 3) The dissertation corresponds to the priority direction of the development of science, approved by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan (indicate the direction)	This work was carried out within the framework of the project for the commercialization of the results of scientific and (or) scientific and technical activities "Opinion Monitoring Information System OMSystem (Opinion monitor system)," 0101-18-GK. The doctoral student developed a method for analyzing social mood using machine and deep learning models, as well as marketing indicators for analyzing the social mood of society. He also conducted a practical experiment on sentiment analysis on the current topic of vaccination against Covid-19 and developed the electronic Social Mood module, which performs additional analysis of data obtained in the monitoring and opinion analysis system OMSystem.

2.	Importance for science	The work **makes** / does not make a significant contribution to science, and its importance is well disclosed / not disclosed	The dissertation has an important contribution to science. It represents a significant step in the development of research, as it describes a newly developed method for analyzing social mood, combining a lexical approach, machine, deep learning models, and marketing measures of social mood. This method makes it possible to significantly improve the accuracy and efficiency of analyzing social mood in the media space. The paper provides not only a theoretical justification for the applicability of the developed method but also its importance is confirmed by experimental studies in analyzing the socio-political topic of vaccination against Covid-19. The results of the study testify to the applicability and effectiveness of the developed method for the analysis and evaluation of social mood in a real context.
3.	The principle of independence	level: 1) **High;** 2) Medium; 3) Low; 4) There is no independence	The doctoral student demonstrated significant autonomy in conducting research and writing this paper. He primarily implemented data processing and the development of an integrated machine learning model using a large number of attribute models and neural networks, as well as a BERT transformer model. Experiments conducted by a doctoral student used trained models to calculate social sentiment scores using marketing metrics. This implies that his research has significant practical significance and can be applied in real experiments to analyze and monitor social mood, for example, on the topic of vaccination against Covid -19, which was given a separate section of this thesis.
4.	The principle of internal unity	4.1 Rationale for the relevance of the dissertation : 1) **Justified;** 2) Partially justified; 3) Not justified.	The dissertation work takes on particular relevance in the context of the modern development of the Internet and the expansion of the media space over the past decades. A huge amount of textual data is observed, which is a valuable source of information about the social mood of a society. The analysis and interpretation of such volumes of data are complex tasks that require the use of modern methods of machine and deep learning, which were done in this work. The special significance of doctoral research extends to the political and social situation in the Republic of Kazakhstan. Understanding the public mood becomes critically important in the context of the country's development and the state's socio-political

		goals. The study results can be used to develop effective policies, programs, and activities that meet the interests and needs of the citizens of Kazakhstan.
	4.2 The content of the dissertation reflects the topic of the dissertation: 1) **Reflects;** 2) Partially reflects; 3) Does not reflect	The content of the dissertation is fully consistent with the topic of the dissertation work and includes the main components: analysis of methods and models of social media monitoring; development of an intelligent system OMSystem and a module that processes and analyzes social media data using machine learning models and neural networks; experimental studies on the application of the developed module for assessing social mood on the example of the important topic of vaccination of the population against coronavirus infection; development of the eSM social sentiment analysis module, which performs additional analysis based on data obtained from the main text database of the OMSystem platform. Thus, the dissertation is a comprehensive research work combining theoretical analysis, development of modules, and experiments in social media analysis.
	4.3. The purpose and objectives correspond to the topic of the dissertation: 1) **Correspond;** 2) Partially correspond; 3) Do not match	The purpose of the dissertation work is to create an effective method for assessing the social mood of society in the media space of the Republic of Kazakhstan. To achieve this goal, the following tasks were set: development of a module capable of analyzing and evaluating social mood using machine learning models, neural networks, and marketing indicators; conducting a practical experiment on using the developed module to analyze the topic of vaccination against Covid-19. The main component of the work is the developed module, which is an intelligent system capable of analyzing and evaluating social mood. The module uses modern machine learning methods, neural networks, and marketing metrics to ensure the accuracy and reliability of the results.
	4.4 All sections and provisions of the dissertation are logically interconnected: 1) **are fully interconnected;** 2) the relationship is partial; 3) there is no relationship	The dissertation sections are connected with each other and have clear and understandable logical transitions. The first section discusses in detail the main aspects and features of information-analytical systems and methods for determining the sentiment of texts, both in domestic and foreign studies. In the second section, the doctoral student describes the newly developed OMSystem platform for monitoring and analyzing the social media space. In

			addition, the construction of the machine and deep learning models for text classification and marketing technologies for analyzing social sentiment, which is used within this platform, are considered. The third section is a comprehensive analysis of the topic of vaccination against Covid-19, which describes in detail the stages of data search, the use of marketing indicators to build analytical results of the experiment in the form of summary tables and graphs, as well as expert conclusions based on the results presented. The fourth section presents the developed electronic Social Mood (eSM) module based on the Django Python framework. This module is an application that parses the data received using the OMSystem platform. The author describes the functionality of the module and its capabilities and explains how it analyzes and evaluates social mood.
		4.5 New solutions proposed by the author (principles, methods) are argued and evaluated in comparison with known solutions: **1) there is a critical analysis;** 2) partial analysis; 3) the analysis is not one's own opinions, but quotes from other authors	In this work, the results of the constructed classification models were evaluated by comparing them with the accepted values of the classification accuracy of machine learning and deep learning algorithms (80% – 90%) and transformer models (91% and higher). Trained models based on the Random Forest algorithm showed the best results, reaching 95% – 99% values, surpassing the standard values for the accuracy of sentiment analysis of texts. Analyzing social mood using marketing technologies is a new approach developed in this work.
5.	The principle of scientific novelty	5.1 Are scientific results and statements new? **1) completely new;** 2) partially new (25-75% are new); 3) not new (less than 25% are new)	The dissertation brings out new scientific provisions, such as a method for analyzing social mood, combining machine learning models to determine the sentiment of texts and marketing indicators for assessing social mood. About ten machine learning and neural network models were implemented as part of the integrated model, including the BERT transformer model. A new sentiment dictionary was also developed, which includes 29654 and 44381 words and word forms of the Kazakh and Russian languages, respectively.
		5.2 Are the conclusions of the dissertation new? **1) completely new;** 2) partially new (25-75% are new); 3) not new (less than 25% are new)	The dissertation work proposes new conclusions on the development of the social mood evaluation module of the OMSystem platform, based on the application of the social mood analysis method using machine learning models, neural networks,

			and marketing indicators of the level of interest in the topic in society, the level of activity of discussing the topic and the level of social mood. The work is supported by experimental results that have been carefully studied and analyzed.
		5.3 Technical, technological, economic, or managerial solutions are new and justified: **1) completely new;** 2) partially new (25-75% are new); 3) not new (less than 25% are new)	In this work, a large number of experiments were carried out aimed at the multiclass classification of the sentiment of texts extracted from the media space of Kazakhstan. The resulting texts have gone through preprocessing stages, in which words are converted to lowercase, and extra words, symbols, punctuation marks, and links are removed. Then, stop words that do not carry a large semantic load were removed. The tf–idf metric and word embedding (FastText , BERT encodings) methods were used to extract the features. At the stage of text classification, various machine learning algorithms were applied, including the Support Vector Machine, K-nearest neighbors, Decision Tree, Random Forest, XGBoost, and others, as well as neural networks, including Deep Neural Network, Convolutional Neural Network, and Recurrent Neural Network. In addition, the BERT transformer model was applied. As a result of the experiments, high classification accuracy values were achieved, ranging from 95% to 98%. These high accuracy scores indicate the models' successful performance and ability to efficiently process and classify large amounts of textual data.
6.	Validity of the main conclusions	All the main conclusions **are based** / not based on scientifically sound evidence or are reasonably well substantiated (for qualitative research and areas of study in the arts and humanities)	The developed method for analyzing social mood, using machine learning methods and marketing technologies, was successfully applied to analyze the current topic of vaccination against Covid-19. The results of the experiments were carefully studied and analyzed in order to understand the reasons for the negative mood in society. The conducted experiment allows a deeper understanding of the public's attitude based on the analysis of the tone and content of the comments.
7.	Basic provisions for defense	The following questions need to be answered for each position separately: 7.1 Is the position proven? **1) proven;** 2) rather proven;	7.1 The developed method of social mood analysis presented in this paper has a wide application for various socio-political topics. In this study, it was successfully applied to the analysis of the topic of vaccination against Covid-19, which provided valuable information about society's opinions and emotional reactions to

		3) rather unproven; 4) not proven 7.2 Is it trivial? 1) yes; **2) no** 7.3 Is it new? **1) yes;** 2) no 7.4 Level to apply: 1) narrow; 2) medium; **3) wide;** 7.5 Is it proven in the article? **1) yes;** 2) no	this important issue. Such analysis is a valuable tool for decision-making and the development of effective public health and social policy strategies. 7.2 Not trivial. 7.3 The provisions put forward for defense are new. 7.4 These methods and models can be applied to various data analysis tasks. 7.5 The results of the dissertation were presented in 12 scientific papers, of which two articles and one chapter were published in journals and book series peer-reviewed in the Scopus database (98 percentile and 71 percentile), four articles in journals recommended by the Committee on quality assurance in the field of education and science of the Ministry of Education and science of the Republic of Kazakhstan, two articles in scientific conferences, peer-reviewed in the Scopus database, and three articles in the proceedings of international conferences.
8.	The principle of certainty Reliability of sources and information provided	8.1 Choice of methodology - justified or methodology described in sufficient detail **1) yes;** 2) no	In the dissertation work, models and methods for analyzing sentiment and social mood are considered in detail. Obtained using a web crawler OMSystem texts went through full stages of preprocessing, feature extraction using metrics and *tf-idf*, and word methods embedding (FastText, BERT encodings). The experiment, which is devoted to the analysis of the reaction of the population to the socio-political measures for vaccination against Covid-19, was also described in detail in the third chapter.
		8.2 The results of the dissertation work were obtained using modern methods of scientific research and methods of processing and interpreting data using computer technologies: **1) yes;** 2) no	In this work, a significant number of software implementations have been performed. The data analysis and processing module was developed in the Python programming language, using a wide range of libraries; in particular, it is important to mention the NLTK (Natural Language Toolkit) library, which is one of the leading libraries in the field of natural language processing. The OMSystem platform and the eSM social sentiment analysis module were developed using the Django framework for Python. Django is a powerful tool that makes it easy to develop web applications, including functionality for query processing, database management, and user interface creation.

			Thus, this work includes a wide range of software implementations based on the Python programming language. The data analysis and processing module was developed using the NLTK library, while the OMSystem platform and the eSM social mood analysis module were built on the Django Python framework, which ensures efficient and convenient development and operation of the system.
		8.3 Theoretical conclusions, models, identified relationships, and patterns are proven and confirmed by experimental research (for areas of training in pedagogical sciences, the results are proven based on a pedagogical experiment): **1) yes;** 2) no	The second and third chapters of this dissertation present a large number of experimental studies both on classifying data with machine learning algorithms and neural networks and using the method of analyzing the social mood of society on the topic of vaccination against Covid-19.
		8.4 Important statements **are supported** / partially confirmed / not supported by references to the relevant and reliable scientific literature	The dissertation contains 114 references to relevant sources.
		8.5 Used literature sources **are sufficient** / not sufficient for a literature review	This number of sources is sufficient for a complete literature review.
9	Principle of practical value	9.1 The dissertation has a theoretical value: **1) yes;** 2) no	The dissertation work is of significant theoretical importance since it includes a detailed study of various aspects. The study examined the architecture and functionality of the OMSystem intellectual system, the main algorithms for machine learning and neural networks, and social analytics marketing indicators, which are important components for assessing social mood. Studying these aspects allows you to better understand the mechanisms of the system's functioning and the basics of its methodology. This theoretical justification is the key to developing and applying various models and algorithms, which in turn ensure the efficiency and accuracy of analyzing social mood in the media space.
		9.2 The dissertation is of practical importance, and there is a high probability of applying the results obtained in practice : **1) yes;** 2) no	The dissertation work has practical application and has already been tested in data analysis on vaccination against Covid-19. The analytical platform and the developed social mood analysis module have great potential for expansion and application in a variety of data analysis tasks.

		9.3 Are the suggestions for practice new? **1) completely new;** 2) partially new (25-75% are new); 3) not new (less than 25% are new)	The proposed analysis of social mood is a new method developed in the course of this research work.
10.	Quality of writing and design	Quality of academic writing: **1) high;** 2) average; 3) below average; 4) low.	The text of the dissertation work has a clear presentation and does not contain serious grammatical and punctuation errors. The clear and concise scientific presentation distinguishes it.

Remarks and suggestions:

1. This work has collected an extensive database, including more than 130 thousand texts. This number labeled by sentiment texts is sufficient for a fundamental analysis of many socio-political topics. However, with the advent of more specific topics, obtaining new data by parsing texts focused on such topics is recommended.

2. The paper carried out a detailed analysis of the topic of vaccination against Covid-19 in 2021, which is an essential experimental result of the developed method for analyzing social mood. In the future, it is recommended to conduct research on new important topics of 2022-2023.

3. Also, in the dissertation, a description was given of using a new sentiment dictionary, which includes 29654 and 44381 words and word forms of the Kazakh and Russian languages, respectively, for labeling texts by sentiment. Although this dictionary covers a large number of words and lexical forms, it is recommended to expand it with new terms to analyze a more significant number of socio-political topics.

The dissertation work of Karyukin Vladislav Igorevich on the topic "**The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan**" complies with the Rules for awarding the degree of Doctor of Philosophy (PhD). Its author Karyukin Vladislav Igorevich deserves to be awarded the Doctor of Philosophy (PhD) degree in the specialty "6D070300 – Information Systems".

Official reviewer:
Doctor of Physical and Mathematical Sciences
Professor of the Department of Information Systems of the
Faculty of Information Technologies
L.N. Gumilyova Eurasian National University

Tussupov J.A.

Қолын куәландырамын
Л.Н. ГУМИЛЕВ АТЫНДАГЫ
ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ
КАДР ҚЫЗМЕТІ

19. 06 2023г.