REVIEW OF THE OFFICIAL REVIEWER

on the dissertation work of Aitim Aigerim Kairatkyzy
on the topic «Models and methods for the automatic processing of unstructured information»,

provided for the degree of Doctor of Philosophy (PhD) in the educational program: 8D06101 – «Clever Systems»

№	Criteria	Eligibility (please select one answer option)	Justification of the official reviewer's position
1	The topic of the thesis (as of the date of its approval) corresponds to the directions of scientific development and/or government programs	1.1 Compliance with priority areas of scientific development or government programs:	The dissertation work corresponds to the objectives of the state programs: 1. National project "Digital Kazakhstan" (Resolution of the Government of the Republic of Kazakhstan dated December 12, 2017 No. 827) - the dissertation contributes to the development of domestic digital technologies and artificial intelligence systems, in particular, in the field of natural language processing (NLP), which meets the goals of digitalization of the public sector and increasing the availability of information in the Kazakh language. 2. The concept of development of the language policy of the Republic of Kazakhstan for 2023-2029 (Decree of the President of the Republic of Kazakhstan dated May 5, 2023 No. 1) - the development of tools for automatic analysis, morphological processing, annotation and generation of Kazakh text contributes to the active promotion and strengthening of the state language in the digital environment. The work directly supports the task of forming a modern language space and digital linguistic technologies. 3. Strategic Development Plan of the Republic of Kazakhstan until 2025 (Decree of the President of the Republic of Kazakhstan dated February 15, 2018 No. 636) - within the framework of the direction "Development of Human Capital and Science", the dissertation contributes to the implementation of scientific developments in practical digital solutions, thereby contributing to increasing the competitiveness of domestic science and education. 4. The Concept of Development of Science of the Republic of Kazakhstan for 2022-2026 (Resolution of the Government of the Republic of Kazakhstan dated October 19, 2021 No. 736) - the study reflects the priority "Development of Information and Communication Technologies and Intelligent Systems", aimed at creating domestic linguistic resources and innovative software solutions, which increases the scientific and technological independence of the country.
		1. The dissertation was completed within the project or target program financed from the state	The dissertation corresponds to the priority area of scientific development "Information and Communication Technologies," approved by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan.

		budget (indicate the name and number of the project or program) 2. The dissertation was completed within the other state program (indicate the name of the program) 3. The thesis corresponds to the priority direction of scientific development, approved by the Higher Scientific and Technical Commission under the Government of the Republic of Kazakhstan (indicate direction)	
2	Importance for science	The work makes/does not make a significant contribution to science and its importance is well-disclosed / not disclosed	The contribution of the work to science is significant, and its importance is well disclosed. The study makes a significant contribution to the development of both domestic and world science in the field of natural language processing, especially in relation to agglutinative and poorly represented languages, such as Kazakh. As part of the work, a complete QNLP library was developed and implemented for the first time, covering the main components of automatic processing of the Kazakh language: morphological analysis, lemmatization, POS tagging, named entity recognition (NER), syntactic parsing and other modules. Original architectural solutions are proposed that adapt modern neural network models (KazBERT + BiLSTM + CRF) to the specific features of the Kazakh language such as vowel harmony, agglutination, free word order and the presence of morphological homonyms. This approach not only expands the scientific understanding of the mechanisms for processing agglutinative languages, but also creates a practical basis for the development of intelligent language technologies with support for the Kazakh language.
3	The principle of self-reliance	Self-reliance level: 1) high; 2) average; 3) low; 4) no independence	The level of independence is high. The dissertation work was completed by the author independently at all stages of scientific research - from formulating the goal, setting tasks and determining the methodological approach to implementing models, conducting experiments, interpreting results and summarizing scientific conclusions. All stages - including the development of the system architecture, collecting and annotating the corpus, implementing algorithms, training models and analyzing their effectiveness, were carried out personally by the author, which indicates a high degree of independence and in-depth study of the topic.

4	The principle of internal unity	4.1 Justification of the relevance of the thesis: 1) justified; 2) partially justified; 3) not justified	The relevance of the dissertation is justified. In the context of the active implementation of key state strategies of the Republic of Kazakhstan, such as "Digital Kazakhstan", "Language Policy Concept for 2023-2029" and "Science Development Concept for 2022-2026", the development of our own intelligent language technologies for the Kazakh language is of strategic importance. Digital transformation of public administration, electronic services, educational platforms, media resources and document flow is impossible without high-quality and adapted solutions in the field of automatic processing of the Kazakh language. The linguistic specificity of the Kazakh language makes the problem particularly relevant: agglutinative nature, vowel harmony, free word order and a high degree of morphological homonymy. These features significantly limit the possibility of direct adaptation of existing Western NLP models, created mainly for analytical and inflectional languages, such as English or German. Accordingly, there is an objective need to develop specialized models, architectures and algorithms that consider the structural, morphological and syntactic features of the Kazakh language. The presented dissertation work meets these challenges and forms a scientifically sound basis for building a full-fledged ecosystem of intelligent linguistic technologies in the state language. Thus, the study makes a significant contribution to the development of national digital infrastructure, scientific linguistics and applied areas of artificial intelligence in the Republic of Kazakhstan.
		4.2 The content of the thesis reflects its topic: 1) reflects; 2) partially reflects; 3) does not reflect	The content of the dissertation reflects the topic of the dissertation, demonstrates a comprehensive and systematic approach to the development of models and methods for automatic processing of unstructured text information in the Kazakh language. The structure of the work is carefully thought out: the chapters are logically structured in accordance with the goal, scientific objectives and main areas of the study, ensuring a consistent disclosure of theoretical foundations, methodological solutions and practical results. Each section complements the previous one, forming a holistic picture of the study - from the analysis of the linguistic features of the Kazakh language and the formulation of problems to the implementation of neural network models and the interpretation of experimental results. This approach contributes to both the scientific novelty and the high applied significance of the work, making its contribution relevant for the development of domestic NLP technologies, the digitalization of the Kazakh language and the implementation of national programs in the field of AI and language policy.
		4.3 The purpose and objectives correspond to the topic of the thesis: 1) correspond; 2) partially correspond;	The purpose and objectives of the study correspond to the topic of the dissertation, since they are aimed at developing effective models and methods for automatic processing of unstructured text information in the Kazakh language. The formulated tasks logically follow from the stated goal and cover all key aspects of the research - from theoretical analysis of the linguistic features of the Kazakh language to the design and implementation of applied software solutions.

		3) do not correspond	Such a structure ensures scientific validity, methodological integrity and completeness of the research, and also contributes to the achievement of both fundamental and practice-oriented results in the field of natural language processing and digitalization of the Kazakh language.
		4.4 All sections and provisions of the dissertation are logically interconnected: 1) completely interconnected; 2) the relationship is partial; 3) there is no relationship	All sections and provisions of the dissertation are logically completely interconnected. All sections, chapters and key provisions of the dissertation research are built in strict accordance with the internal logic of scientific analysis and are subordinated to a single goal - a consistent and reasonable solution to the scientific problems. The structure of the work reflects the logical dynamics of the research: from formulating the problem and analyzing theoretical and applied aspects - to designing, implementing and empirically evaluating the proposed solutions. This approach ensures the integrity, coherence and methodological rigor of the presentation, allowing you to trace the development of scientific thought at all stages - from formulating hypotheses to obtaining practical results and formulating conclusions.
		4.5 The new solutions (principles, methods) proposed by the author are argued and evaluated in comparison with known solutions: 1) there is a critical analysis; 2) partial analysis; 3) the analysis does not represent one's own opinions, but quotes from other authors	The dissertation proposes original principles, methods and architectures aimed at automatic processing of unstructured text information in the Kazakh language, the effectiveness of which has been confirmed both theoretically and experimentally, and includes a critical analysis. The author has developed and implemented a combined architecture of KazBERT + BiLSTM + CRF, specially adapted to the linguistic features of the Kazakh language, including agglutination, vowel harmony, free word order and morphological homonymy, which has significantly increased the accuracy of morphological and syntactic markup. In addition, a comprehensive QNLP framework has been created that combines morphoanalysis, lemmatization, POS tagging, NER and syntactic analysis modules into a single system. All solutions have been experimentally tested on our own corpus of the Kazakh language and compared with leading tools (Stanza, UDPipe, Trankit, spaCy), showing an increase in the accuracy of POS tagging by 5-12%, an improvement in the quality of NER and more accurate morphemic division. This confirms both the scientific novelty of the developments and their practical applicability in the tasks of digitalization of the Kazakh language.
5	The principle of scientific novelty	5.1 Are the scientific results and provisions new? 1) completely new; 2) partially new (25–75% are new); 3) not new (less than 25% are new)	All the main provisions and scientific results presented in the dissertation are completely new, original and reflect the personal contribution of the author to the development of natural language processing methods in relation to the Kazakh language. They correspond to the current level of research in the field of artificial intelligence, computational linguistics, and technologies for processing agglutinative languages. The key scientific results obtained during the study include: development and implementation of the original KazBERT + BiLSTM + CRF architecture for the tasks of morphological analysis, POS tagging and named entity recognition (NER), adapted to the features of the Kazakh language; creation

			of the QNLP framework - the first comprehensive open-source system integrating modules of morphological analysis, lemmatization, POS tagging, NER and syntactic analysis for the Kazakh language; conducting a systemic comparative analysis of modern NLP tools with experimental validation confirming the advantage of the proposed solutions. The reliability and practical significance of the results are confirmed by their testing in the form of 13 publications, reports at scientific conferences, two acts of implementation and two author's certificates, which indicates a high scientific level and applied value of the work performed.
		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 conclusions of the dissertation are completely new, logically consistent, and completely follow from the goals and objectives of the study. They reflect the results of the development and experimental evaluation of original models and methods for automatic processing of Kazakh-language texts that have not previously been presented in the scientific literature. Scientific novelty lies in the integration of the adapted KazBERT + BiLSTM + CRF architecture, capable of considering the linguistic features of the Kazakh language, formalizing the rules of word formation and morphemic analysis, as well as creating and annotating our own text corpus covering the key linguistic phenomena of the Kazakh language. The findings have both theoretical significance - contributing to the development of linguistic models for agglutinative languages, and high practical value, providing a basis for the creation of intelligent language technologies. The reliability and significance of the findings are confirmed by the results of comparative experiments, publications in scientific journals, participation in international and national conferences, as well as testing of developments on practical platforms.
		5.3 Technical, technological, economic, or managerial decisions are 1) new and justified; 2) completely new; 3) partially new (25–75% are new); 4) not new (less than 25% are new)	The technical and technological solutions presented in the dissertation are entirely novel, as evidenced by the development and implementation of a comprehensive system for the automatic processing of unstructured information in the Kazakh language. As part of the study, an original library QNLP was created, which includes modules for morphological analysis, part-of-speech and named entity tagging, syntactic and semantic analysis, all specifically adapted to the agglutinative nature of the Kazakh language. The developed ontology of Kazakh morphology and the lexicographic database supporting automatic word form generation and lemmatization, along with the construction of a proprietary annotated corpus of Kazakh texts, confirm the high level of scientific and practical novelty of this work, which has no direct equivalents in existing domestic or international research.
6	Validity of the main conclusions	All key findings are based / are not based on scientifically sound evidence or are reasonably well funded (for qualitative research and areas of training in the arts	The conclusions formulated in the dissertation are scientifically based, new and logically follow from the goals and objectives of the study. The main conclusions made based on publication in scientific journals of databases indexed in Scopus: Aitim A.K., Satybaldiyeva R.Zh. A comparison of Kazakh language processing models for improving semantic search results, Eastern-European Journal of Enterprise Technologies, (Q3) a high-impact scientific journal with a citation impact rating of 2.0 and a percentile

		and humanities)	of 46, 1(2 (133), 66–75, 2025. https://doi.org/10.15587/1729-4061.2025.315954 Also, in recommended publications by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Education and Science of the Republic of Kazakhstan: 1. Aitim A.K., Satybaldiyeva R.Zh., Linguistic ontology as means of modeling of a coherent text, Bulletin of Abai KazNPU. Series of Physical and mathematical sciences. 3 (Sep. 2022), https://doi.org/10.51889/3879.2022.77.24.017 2. Aitim A.K., Developing methods for automatic processing systems of Kazakh language, Bulletin of KazATC 133 (4), 2024, ISSN 1609-1817, ISSN Online 2790-5802, https://doi.org/10.52167/1609-1817 3. Aitim A.K., Satybaldiyeva R.Zh., A systematic review of existing tools to automated processing systems for Kazakh language. Bulletin of Abai KazNPU. Series of Physical and mathematical sciences. 87, 3 (Sep. 2024), 106–122, https://doi.org/10.51889/2959-5894.2024.87.3.009 4. Aitim A.K., Satybaldiyeva R.Zh., Building methods and models for automatic processing systems of Kazakh language, Bulletin of KazATC No. 2-137-2025, https://doi.org/10.52167/1609-1817-2024-133-4-254-265
7	Main provisions submitted for defense	The following questions must be answered for each provision separately: 7.1 Has the position been proven? 1) proven**;** 2) rather proven; 3) rather not proven; 4) not proven 7.2 Is it trivial? 1) yes; 2) no 7.3 Is it new? 1) yes; 2) no 7.4 Application level: 1) narrow; 2) medium; 3) wide 7.5 Is it proven in the	7.1 The main provisions submitted for defense are fully proven. 7.2 The main provisions submitted for defense contain new theoretical knowledge, methods of solution and implementation and are not trivial. 7.3 The provisions and scientific achievements are new, which is confirmed by the high percentage of originality 7.4 The models and software solutions developed in the dissertation have a wide level of applicability, covering both scientific and applied areas. In the scientific context, the results obtained can be used for further development of methods for processing agglutinative languages, formalization of morphological and syntactic structures and creation of specialized language resources. In the applied sphere, the proposed algorithms and the QNLP library open up wide opportunities for implementation in intelligent text analysis systems, educational online platforms, search engines, voice interfaces and digital assistants with support for the Kazakh language. Due to adaptation to the grammatical features of the Kazakh language, as well as modular architecture and open structure, the developed solutions have the potential to be scaled to other Turkic languages, which significantly expands their scope of application and contributes to the development of interlingual NLP technologies in the region. 7.5 Proved in the article The provisions of the dissertation research are proven by the studies conducted. The main provisions submitted for defense are presented in 13 publications of the applicant: The main results obtained in the course of the dissertation work were published in thirteen

		article? 1) yes; 2) no	printed works, including 4 articles in publications recommended by the Committee for Control in the Sphere of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, as well as 2 articles in the journal Eastern-European Journal of Enterprise Technologies (Q3), indexed in the Scopus database, with a cite score of 2.0 and a percentile of 46. In addition, 4 articles were published in the materials of international conferences, including one scientific article with a cite score of 4.5 and a percentile of 69 (Q2) and 3 articles in other republican publications.
8	The principle of reliability. Reliability of sources and information provided	8.1 The choice of methodology – is the methodology justified or is it described in sufficient detail? 1) yes; 2) no	The research methodology is carefully substantiated and described in detail: the choice of architectures, methods of data preprocessing, annotation and evaluation of models is dictated by both the linguistic features of the Kazakh language and the scientific objectives set in the dissertation. All stages - from corpus formation and linguistic analysis to model design and experimental validation - are presented in a logical sequence and are accompanied by a reasoned choice of the methods and tools used. This approach ensures both the transparency and reproducibility of the study, as well as its scientific validity, allowing the use of the obtained results in subsequent developments and theoretical research in related fields.
		8.2 The results of the thesis were obtained using modern methods of scientific research and techniques for processing and interpreting data using computer technologies: 1) yes; 2) no	The results of the dissertation were obtained using modern scientific methods covering the areas of machine learning, neural network architectures and natural language processing technologies. In the process of developing models and analyzing data, reliable statistical methods for quality assessment were used, ensuring objective validation of the results, as well as specialized software for preprocessing, annotation and visualization of large volumes of texts in the Kazakh language. All stages of the research - from collecting and preparing the corpus to building and testing the models - were implemented using current computer technologies and tools, which guaranteed high accuracy, reproducibility and practical significance of the obtained scientific results.
		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	Theoretical conclusions, developed models, as well as the identified relationships and patterns within the framework of the dissertation research are proven and confirmed by the results of experimental studies. All proposed solutions - from morphological analysis algorithms to a comprehensive QNLP framework - have been practically tested on a specially formed and representative corpus of Kazakh-language texts. The effectiveness and correctness of the proposed models were validated using generally accepted quality metrics (such as Accuracy, Precision, Recall, F1-score) and compared with the results of existing NLP systems. This ensures a high degree of reliability, reproducibility and scientific validity of the obtained data, confirming the practical value and practical applicability of the developments.

		8.4 Important statements are supported / partially supported / not supported by references to current and reliable scientific literature	Important statements are supported by references to relevant and reliable scientific literature.
		8.5 The literature sources used are sufficient / are not sufficient for the literature review	The literature sources used in the dissertation are sufficient in volume and representative in content, covering a wide range of modern scientific works in the field of natural language processing, machine learning, agglutinative linguistics and digital processing of the Kazakh language. The literature review analyzes both fundamental and classical approaches, as well as advanced research of recent years, including publications in highly rated international journals and materials of leading scientific conferences. This approach ensures a comprehensive theoretical basis for the study, reflects its relevance and integration into the modern scientific context.
9	Principle of practical value	9.1 The dissertation has theoretical significance: 1) yes; 2) no	The dissertation has a significant theoretical contribution, since it develops and substantiates new approaches to the automatic processing of agglutinative languages using the Kazakh language as an example. Original models of morphological analysis, POS tagging and semantic markup adapted to the structural features of the Kazakh language are presented. The proposed theoretical provisions contribute to the expansion of the modern scientific understanding of the formalization of grammatical and morphological processes in agglutinative languages and contribute to the development of computational linguistics. In addition, the dissertation forms a methodological basis for further research in the field of automatic text processing and the creation of language technologies for low-resource languages, including other Turkic languages.
		9.2 The thesis has practical significance and there is a high probability of applying the obtained results in practice: 1) yes; 2) no	The dissertation has a pronounced practical significance, since within its framework a full-featured QNLP software framework for automatic text processing in the Kazakh language was developed. The system includes modules of morphological analysis, POS tagging, named entity recognition (NER) and syntactic analysis adapted to the linguistic features of the agglutinative structure of the Kazakh language. The results obtained have high applied value and can be directly implemented in a wide range of IT solutions - from electronic document management systems and digital educational platforms to voice interfaces, chatbots and intelligent assistants in the Kazakh language. The open architecture of the library, modularity and the ability to scale to other Turkic languages confirm the potential for integrating developments into both government and commercial projects aimed at digitalization and expanding the use of the Kazakh language in the technological environment.

		9.3 Are the suggestions for practice new? 1) completely new; 2) partially new (25–75% are new); 3) not new	The proposals formulated in the dissertation for practical application are completely new, original and focused on the real needs of digital processing of the Kazakh language. They are based on the author's developments, specially adapted to the morphological and syntactic features of the Kazakh language. For the first time, a comprehensive approach to building a QNLP language framework has been proposed, ensuring the integration of morphological analysis, syntactic and semantic processing modules into a single system. An essential part of the practical recommendations is the development of automated tools for pre-notation and visual markup of texts, which opens up broad prospects for creating scalable and high-quality language resources. The presented solutions can be effectively applied when implementing NLP technologies in government information systems, educational platforms, intelligent search engines and commercial digital services focused on the Kazakh language.
10.	Quality of writing and design	Quality of academic writing: 1) high; 2) average; 3) below average; 4) low.	The quality of academic writing in the dissertation is high, the text is characterized by a logical structure, consistent presentation of material and precision of wording. The presentation is conducted in a scientific style with the correct use of terms from the fields of linguistics, computer science and artificial intelligence. All theoretical provisions are supported by arguments, and the conclusions are substantiated by the results of the research. A clear structure and clarity of presentation contribute to the effective disclosure of the content and demonstrate a high level of academic culture, ensuring scientific credibility and ease of perception of the presented material.
11.	Comments and suggestions on the dissertation		*The literature review is dominated by English-language works, which, on the one hand, reflects the international level of research, but on the other hand, does not sufficiently take into account the local context. It would be necessary to more widely involve studies carried out in the Kazakh and Russian languages in the field of linguistics, lexicography and Kazakh morphology, including the work of domestic researchers. The above comments are advisory in nature and do not have a significant impact on the scientific significance and applied value of the results obtained.*
12.	The scientific level of the doctoral candidate's articles on the research topic (in the case of defending the dissertation in the form of a series of articles, the official reviewers will comment on the scientific level of each		The scientific level of the applicant’s publications fully complies with the requirements for PhD dissertation.

	article by the doctoral candidate on the research topic).		
13.	The decision of the official reviewer (according to paragraph 28 of the current Model Regulation)		To petition the Committee for the award of the degree of Doctor of Philosophy (PhD)

Conclusion

The presented dissertation for the degree of Doctor of Philosophy (PhD) by **Aitim Aigerim Kairatkyzy** on the topic **«Models and methods for the automatic processing of unstructured information»**, meets the requirements of the Rules, the author deserves to petition the Committee for the award of the degree of Doctor of Philosophy (PhD).

Official reviewer:
PhD, professor, deputy general director of the
Institute of Information and Computing Technologies **Mamyrbaev Orken Zhumazhanovich**

Copies of reviews from official reviewers are given to the doctoral student no later than 5 (five) working days before the defense of the dissertation.

ОТЗЫВ ОФИЦИАЛЬНОГО РЕЦЕНЗЕНТА

на диссертационную работу Әйтім Әйгерім Қайратқызы на тему «Models and methods for the automatic processing of unstructured information» («Модели и методы автоматической обработки неструктурированной информации»),

предоставленную на соискание степени доктора философии (PhD) по образовательной программе 8D06101 – «Интеллектуальные системы»

п/п	Критерии	Соответствие критериям (подчеркнуть один из вариантов ответа)	Обоснование позиции официального рецензента (замечания выделить курсивом)
1	Тема диссертации (на дату ее утверждения) соответствует направлениям развития науки и/или государственным программам	1.1 Соответствие приоритетным направлениям развития науки или государственным программам:	Диссертационная работа соответстует задачам государственных программ: 1. Национальный проект «Цифровой Казахстан» (постановление Правительства РК от 12 декабря 2017 года № 827) - диссертация способствует развитию отечественных цифровых технологий и систем искусственного интеллекта, в частности, в области обработки естественного языка (NLP), что отвечает целям цифровизации государственного сектора и повышения доступности информации на казахском языке. 2. Концепция развития языковой политики Республики Казахстан на 2023–2029 годы (указ Президента РК от 5 мая 2023 года № 1) - разработка инструментов для автоматического анализа, морфологической обработки, аннотации и генерации казахского текста способствует активному продвижению и укреплению государственного языка в цифровой среде. Работа напрямую поддерживает задачу формирования современного языкового пространства и цифровых лингвистических технологий. 3. Стратегический план развития Республики Казахстан до 2025 года (указ Президента РК от 15 февраля 2018 года № 636) – в рамках направления «Развитие человеческого капитала и науки» диссертация способствует внедрению научных разработок в практические цифровые решения, тем самым обеспечивая вклад в повышение конкурентоспособности отечественной науки и образования. 4. Концепция развития науки Республики Казахстан на 2022–2026 годы (постановление Правительства РК от 19 октября 2021 года № 736) - исследование отражает приоритет «Развитие информационно-коммуникационных технологий и интеллектуальных систем», направлено на создание отечественных лингвистических ресурсов и инновационных

			программных решений, что повышает научную и технологическую независимость страны.
		1) Диссертация выполнена в рамках проекта или целевой программы, финансируемого(ой) из Государственного бюджета (указать название и номер проекта или программы) 2) Диссертация выполнена в рамках другой государственной программы (указать название программы) 3) Диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики	Диссертация соответствует приоритетному направлению развития науки «Информационные, коммуникационные технологии», утвержденному высшей научно-технической комиссией при Правительстве РК.

		Казахстан (указать направление)	
2	Важность для науки	Работа вносит/не вносит существенный вклад в науку, а ее важность хорошо раскрыта/не раскрыта	Вклад работы в науку является существенным, а важность раскрыта хорошо. Исследование делает весомый вклад в развитие как отечественной, так и мировой науки в области обработки естественного языка, особенно применительно к агглютинативным и слабо представленным языкам, таким как казахский. В рамках работы впервые разработана и реализована целостная библиотека QNLP, охватывающая основные компоненты автоматической обработки казахского языка: морфологический анализ, лемматизацию, POS-теггинг, распознавание именованных сущностей (NER), синтаксический разбор и другие модули. Предложены оригинальные архитектурные решения, адаптирующие современные нейросетевые модели (KazBERT + BiLSTM + CRF) к специфическим особенностям казахского языка таким как сингармонизм, агглютинация, свободный порядок слов и наличие морфологических омонимий. Такой подход не только расширяет научное понимание механизмов обработки агглютинативных языков, но и создает практическую основу для разработки интеллектуальных языковых технологий с поддержкой казахского языка.
3	Принцип самостоятельности	Уровень самостоятельности: 1) Высокий; 2) Средний; 3) Низкий; Самостоятельности нет	Уровень самостоятельности – высокий. Диссертационная работа выполнена автором самостоятельно на всех этапах научного исследования - от формулирования цели, постановки задач и определения методологического подхода до реализации моделей, проведения экспериментов, интерпретации результатов и обобщения научных выводов. Все этапы - включая разработку архитектуры системы, сбор и аннотирование корпуса, реализацию алгоритмов, обучение моделей и анализ их эффективности, осуществлены лично автором, что свидетельствует о высокой степени самостоятельности и глубокой проработке темы.
4	Принцип внутреннего единства	4.1 Обоснование актуальности диссертации: 1) Обоснована; 2) Частично обоснована; 3) Не обоснована.	Обоснована актуальность диссертации. В условиях активной реализации ключевых государственных стратегий Республики Казахстан, таких как «Цифровой Казахстан», «Концепция языковой политики на 2023–2029 годы» и «Концепция развития науки на 2022–2026 годы» разработка собственных интеллектуальных языковых технологий для казахского языка приобретает стратегически важное значение. Цифровая трансформация государственного управления, электронных услуг, образовательных платформ, медиа-ресурсов и документооборота невозможна без качественных и адаптированных решений в области автоматической

			обработки казахского языка. Особую актуальность проблеме придаёт лингвистическая специфика казахского языка: агглютинативная природа, сингармонизм, свободный порядок слов и высокая степень морфологической омонимии. Эти особенности существенно ограничивают возможность прямой адаптации существующих западных NLP-моделей, создаваемых в основном для аналитических и флективных языков, таких как английский или немецкий. Соответственно, возникает объективная необходимость в разработке специализированных моделей, архитектур и алгоритмов, учитывающих структурные, морфологические и синтаксические особенности казахского языка. Представленная диссертационная работа отвечает этим вызовам и формирует научно обоснованную основу для построения полноценной экосистемы интеллектуальных лингвистических технологий на государственном языке. Таким образом, исследование вносит существенный вклад в развитие национальной цифровой инфраструктуры, научной лингвистики и прикладных направлений искусственного интеллекта в Республике Казахстан.
		4.2 Содержание диссертации отражает тему диссертации: 1) Отражает; 2) Частично отражает; 3) Не отражает	Содержание диссертации отражает тему диссертации, демонстрирует комплексный и системный подход к разработке моделей и методов автоматической обработки неструктурированной текстовой информации на казахском языке. Структура работы тщательно продумана: главы логично выстроены в соответствии с целью, научными задачами и основными направлениями исследования, обеспечивая последовательное раскрытие теоретических основ, методологических решений и практических результатов. Каждый раздел дополняет предыдущий, формируя целостную картину проведённого исследования - от анализа лингвистических особенностей казахского языка и постановки задач до реализации нейросетевых моделей и интерпретации экспериментальных результатов. Такой подход способствует как научной новизне, так и высокой прикладной значимости работы, делая её вклад актуальным для развития отечественных NLP-технологий, цифровизации казахского языка и реализации национальных программ в области ИИ и языковой политики.
		4.3. Цель и задачи соответствуют теме диссертации: 1) соответствуют; 2) частично соответствуют; 3) не соответствуют	Цель и задачи исследования соответствуют теме диссертации, поскольку направлены на разработку эффективных моделей и методов автоматической обработки неструктурированной текстовой информации на казахском языке. Сформулированные задачи логично вытекают из поставленной цели и охватывают все ключевые аспекты исследования - от теоретического анализа лингвистических особенностей казахского языка до проектирования и реализации прикладных программных решений.

			Такое построение обеспечивает научную обоснованность, методологическую целостность и полноту исследования, а также способствует достижению как фундаментальных, так и практико-ориентированных результатов в области обработки естественного языка и цифровизации казахского языка.
		4.4. Все разделы и положения диссертации логически взаимосвязаны: 1) полностью взаимосвязаны; 2) взаимосвязь частичная; взаимосвязь отсутствует	Все разделы и положения диссертации логически полностью взаимосвязаны. Все разделы, главы и ключевые положения диссертационного исследования выстроены в строгом соответствии с внутренней логикой научного анализа и подчинены единой цели - последовательному и обоснованному решению поставленных научных задач. Структура работы отражает логическую динамику исследования: от формулирования проблемы и анализа теоретических и прикладных аспектов - к проектированию, реализации и эмпирической оценке предложенных решений. Такой подход обеспечивает целостность, связность и методологическую строгость изложения, позволяя проследить развитие научной мысли на всех этапах - от постановки гипотез до получения практических результатов и формулирования выводов.
		4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями: 1) критический анализ есть; 2) анализ частичный; анализ представляет собой не собственные мнения, а цитаты других авторов	В диссертации предложены оригинальные принципы, методы и архитектуры, направленные на автоматическую обработку неструктурированной текстовой информации на казахском языке, эффективность которых подтверждена как теоретически, так и экспериментально, есть критический анализ. Автор разработал и реализовал комбинированную архитектуру KazBERT + BiLSTM + CRF, специально адаптированную под лингвистические особенности казахского языка, включая агглютинацию, сингармонизм, свободный порядок слов и морфологическую омонимию, что позволило существенно повысить точность морфологической и синтаксической разметки. Кроме того, создан комплексный фреймворк QNLP, объединяющий в единую систему модули морфоанализа, лемматизации, POS-теггинга, NER и синтаксического анализа. Все решения прошли экспериментальную проверку на собственном корпусе казахского языка и были сопоставлены с ведущими инструментами (Stanza, UDPipe, Trankit, spaCy), показав прирост точности POS-теггинга на 5–12%, улучшение качества NER и более точное морфемное членение. Это подтверждает как научную новизну разработок, так и их практическую применимость в задачах цифровизации казахского языка.

5	Принцип научной новизны	5.1 Научные результаты и положения являются новыми? 1) полностью новые; 2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее25%)	Все основные положения и научные результаты, представленные в диссертационной работе, являются полностью новыми, оригинальными и отражают личный вклад автора в развитие методов обработки естественного языка применительно к казахскому языку. Они соответствуют современному уровню исследований в области искусственного интеллекта, вычислительной лингвистики и технологий обработки агглютинативных языков. К числу ключевых научных результатов, полученных в ходе исследования, относятся: разработка и реализация оригинальной архитектуры KazBERT + BiLSTM + CRF для задач морфологического анализа, POS-теггинга и распознавания именованных сущностей (NER), адаптированной под особенности казахского языка; создание фреймворка QNLP - первой комплексной open-source системы, интегрирующей модули морфологического анализа, лемматизации, POS-теггинга, NER и синтаксического анализа для казахского языка; проведение системного сравнительного анализа современных NLP-средств с экспериментальной валидацией, подтверждающей преимущество предложенных решений. Достоверность и практическая значимость результатов подтверждены их апробацией в форме 13 публикаций, докладов на научных конференциях, двух актов внедрения и двух авторских свидетельств, что свидетельствует о высоком научном уровне и прикладной ценности выполненной работы.
		5.2 Выводы диссертации являются новыми? 1) полностью новые; 2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее25%)	Выводы диссертационной работы являются полностью новыми, логически последовательными и полностью вытекают из целей и задач исследования. Они отражают результаты разработки и экспериментальной оценки оригинальных моделей и методов автоматической обработки казахскоязычных текстов, ранее не представленных в научной литературе. Научная новизна заключается, в частности, в интеграции адаптированной архитектуры KazBERT + BiLSTM + CRF, способной учитывать лингвистические особенности казахского языка, формализации правил словообразования и морфемного анализа, а также создании и аннотировании собственного текстового корпуса, охватывающего ключевые лингвистические феномены казахского языка. Полученные выводы обладают как теоретической значимостью - способствуя развитию лингвистических моделей для агглютинативных языков, так и высокой практической ценностью, обеспечивая основу для создания интеллектуальных языковых технологий. Достоверность и значимость выводов подтверждаются результатами сравнительных экспериментов, публикациями в научных изданиях, участием в международных и национальных конференциях, а также апробацией разработок на практических платформах.

		5.3 Технические, технологические, экономические или управленческие решения являются новыми и обоснованными: 1) полностью новые; 2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее 25%)	Представленные в диссертационной работе технические и технологические решения являются полностью новыми, что подтверждается разработкой и реализацией первой в своём роде комплексной системы автоматической обработки неструктурированной информации на казахском языке. В рамках исследования создана уникальная библиотека QNLP, включающая модули морфологического анализа, частеречной и именованной разметки, синтаксического и семантического анализа, специально адаптированные под структурные и грамматические особенности агглютинативной природы казахского языка. Разработанная онтология казахской морфологии и лексикографическая база, обеспечивающая автоматическую генерацию словоформ, лемматизацию и морфемный разбор, а также формирование собственного полноценно размеченного корпуса казахских текстов, подтверждают высокий научный и прикладной уровень выполненной работы. Эти решения не имеют прямых аналогов среди существующих отечественных и международных разработок, что подчёркивает их инновационный характер и значимость для дальнейшего развития технологий обработки естественного языка для казахского и других агглютинативных языков.
6	Обоснованность основных выводов	Все основные выводы основаны/не основаны на весомых с научной точки зрения доказательствах либо достаточно хорошо обоснованы	Выводы, сформулированные в диссертационной работе, являются научно обоснованными, новыми и логически вытекающими из целей и задач исследования. Основные выводы, сделанные на основе публикации в научных журналах баз данных, индексируемых в Scopus: Aitim A.K., Satybaldiyeva R.Zh. A comparison of Kazakh language processing models for improving semantic search results, Eastern-European Journal of Enterprise Technologies, (Q3) высокоэффективный научный журнал с рейтингом цитируемости 2,0 и процентилем 46, 1(2 (133), 66–75, 2025. https://doi.org/10.15587/1729-4061.2025.315954 А также рекомендованных в изданиях Комитетом по обеспечению качества в сфере науки и высшего образования МНВО РК: 1. Aitim A.K., Satybaldiyeva R.Zh., Linguistic ontology as means of modeling of a coherent text, Bulletin of Abai KazNPU. Series of Physical and mathematical sciences. 3 (Sep. 2022), https://doi.org/10.51889/3879.2022.77.24.017 2. Aitim A.K., Developing methods for automatic processing systems of Kazakh language, Bulletin of KazATC 133 (4), 2024, ISSN 1609-1817, ISSN Online 2790-5802, https://doi.org/10.52167/1609-1817 3. Aitim A.K., Satybaldiyeva R.Zh., A systematic review of existing tools to

			automated processing systems for Kazakh language. Bulletin of Abai KazNPU. Series of Physical and mathematical sciences. 87, 3 (Sep. 2024), 106–122, https://doi.org/10.51889/2959-5894.2024.87.3.009 4. Aitim A.K., Satybaldiyeva R.Zh., Building methods and models for automatic processing systems of Kazakh language, Bulletin of KazATC №2-137-2025, https://doi.org/10.52167/1609-1817-2024-133-4-254-265
7	Основные положения, выносимые на защиту	Необходимо ответить на следующие вопросы по каждому положению в отдельности: 7.1 Доказано ли положение? 1) доказано; 2) скорее доказано; 3) скорее не доказано; не доказано 7.2 Является ли тривиальным? 1) да; 2) нет 7.3 Является ли новым? 1) да; 2) нет 7.4 Уровень для применения: 1) узкий; 2) средний; 3) широкий 7.5 Доказано ли в статье? 1) да; 2) нет	7.1 Основные положения выносимые на защиту полностью доказаны. 7.2 Основные положения, выносимые на защиту, содержат новые теоретические знания, способы решения и реализации и не являются тривиальными. 7.3 Положения и научные достижения являются новыми, что подтверждается высоким процентом оригинальности 7.4 Разработанные в диссертационной работе модели и программные решения обладают широким уровнем применимости, охватывая как научные, так и прикладные области. В научном контексте полученные результаты могут быть использованы для дальнейшего развития методов обработки агглютинативных языков, формализации морфологических и синтаксических структур и создания специализированных языковых ресурсов. В прикладной сфере предложенные алгоритмы и библиотека QNLP открывают широкие возможности для внедрения в интеллектуальные системы анализа текстов, образовательные онлайн-платформы, поисковые механизмы, голосовые интерфейсы и цифровых ассистентов с поддержкой казахского языка. Благодаря адаптации к грамматическим особенностям казахского языка, а также модульной архитектуре и открытой структуре, разработанные решения обладают потенциалом масштабирования на другие тюркские языки, что существенно расширяет их область применения и способствует развитию межъязыковых NLP-технологий в регионе. 7.5 Доказаны в статье Положения диссертационного исследования доказаны проведенными исследованиями. Основные положение, выносимые на защиту, представлены в 13 публикациях соискателя: Основные результаты, полученные в ходе выполнения диссертационной работы, были опубликованы в тринадцати печатных работах, включая 4 статьи в изданиях, рекомендованных Комитетом по контролю в сфере образования и науки Министерства образования и науки Республики Казахстан, а также 2 статьи в журнале Eastern-European Journal of Enterprise Technologies (Q3), индексируемом в базе Scopus, с cite score 2.0 и перцентилем 46. Кроме того, опубликовано 4 статей в материалах

			международных конференций, среди которых одна научная статья с cite score 4.5 и перцентилем 69 (Q2) и 3 статьи в других республиканских изданиях.
8	Принцип достоверности Достоверность источников и предоставляемой информации	8.1 Выбор методологии - обоснован или методология достаточно подробно описана 1) да; 2) нет	Методология исследования тщательно обоснована и детально описана: выбор архитектур, методов предобработки данных, аннотации и оценки моделей продиктован как лингвистическими особенностями казахского языка, так и поставленными в диссертации научными задачами. Все этапы - от формирования корпуса и лингвистического анализа до проектирования моделей и экспериментальной валидации - представлены в логической последовательности и сопровождаются аргументированным выбором применяемых методов и инструментов. Такой подход обеспечивает как прозрачность и воспроизводимость исследования, так и его научную состоятельность, позволяя использовать полученные результаты в последующих разработках и теоретических изысканиях в смежных областях.
		8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий: 1) да; 2) нет	Результаты диссертационной работы получены с применением современных научных методов, охватывающих области машинного обучения, нейросетевых архитектур и технологий обработки естественного языка. В процессе разработки моделей и анализа данных использовались надёжные статистические методы оценки качества, обеспечивающие объективную валидацию результатов, а также специализированные программные средства для предобработки, аннотации и визуализации больших объёмов текстов на казахском языке. Все этапы исследования - от сбора и подготовки корпуса до построения и тестирования моделей - были реализованы с использованием актуальных компьютерных технологий и инструментов, что гарантировало высокую точность, воспроизводимость и прикладную значимость полученных научных результатов.
		8.3 Теоретические выводы, модели, выявленные взаимосвязи и закономерности доказаны и подтверждены экспериментальны	Теоретические выводы, разработанные модели, а также выявленные взаимосвязи и закономерности в рамках диссертационного исследования доказаны и подтверждены результатами экспериментальных исследований. Все предложенные решения - от алгоритмов морфологического анализа до комплексного фреймворка QNLP прошли практическую проверку на специально сформированном и репрезентативном корпусе казахскоязычных текстов. Эффективность и корректность предложенных моделей были валидированы с использованием общепринятых метрик качества (таких как

		м исследованием (для направлений подготовки по педагогическим наукам результаты доказаны на основе педагогического эксперимента): 1) да; 2) нет	Accuracy, Precision, Recall, F1-score) и сравнены с результатами существующих NLP-систем. Это обеспечивает высокую степень достоверности, воспроизводимости и научной обоснованности полученных данных, подтверждая практическую ценность и прикладную применимость разработок.
		8.4 Важные утверждения подтверждены/частично подтверждены/не подтверждены ссылками на актуальную и достоверную научную литературу	Важные утверждения подтверждены ссылками на актуальную и достоверную научную литературу
		8.5 Использованные источники литературы достаточны/не достаточны для литературного обзора	Использованные в диссертационной работе источники литературы являются достаточными по объёму и репрезентативными по содержанию, охватывая широкий спектр современных научных трудов в области обработки естественного языка, машинного обучения, агглютинативной лингвистики и цифровой обработки казахского языка. В литературном обзоре проанализированы как фундаментальные и классические подходы, так и передовые исследования последних лет, включая публикации в высокорейтинговых международных журналах и материалах ведущих научных конференций. Такой подход обеспечивает всестороннюю теоретическую обоснованность исследования, отражает его актуальность и интеграцию в современный научный контекст.

9	Принцип практической ценности	9.1 Диссертация имеет теоретическое значение: 1) да; 2) нет	Диссертация обладает значительным теоретическим вкладом, поскольку в ней разработаны и обоснованы новые подходы к автоматической обработке агглютинативных языков на примере казахского. Представлены оригинальные модели морфологического анализа, POS-теггинга и семантической разметки, адаптированные к особенностям структуры казахского языка. Предложенные теоретические положения способствуют расширению современного научного представления о формализации грамматических и морфологических процессов в агглютинативных языках и вносят вклад в развитие вычислительной лингвистики. Кроме того, диссертация формирует методологическую основу для дальнейших исследований в области автоматической обработки текстов и создания языковых технологий для малоресурсных языков, включая другие тюркские языки.
		9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике: 1) да; 2) нет	Диссертация обладает выраженной практической значимостью, поскольку в её рамках разработан полнофункциональный программный фреймворк QNLP для автоматической обработки текстов на казахском языке. Система включает модули морфологического анализа, POS-теггинга, распознавания именованных сущностей (NER) и синтаксического анализа, адаптированные к лингвистическим особенностям агглютинативной структуры казахского языка. Полученные результаты имеют высокую прикладную ценность и могут быть непосредственно внедрены в широкий спектр ИТ-решений - от систем электронного документооборота и цифровых образовательных платформ до голосовых интерфейсов, чат-ботов и интеллектуальных ассистентов на казахском языке. Открытая архитектура библиотеки, модульность и возможность масштабирования на другие тюркские языки подтверждают потенциал интеграции разработок как в государственные, так и в коммерческие проекты, направленные на цифровизацию и расширение использования казахского языка в технологической среде.
		9.3 Предложения для практики являются новыми? 1) полностью новые; 2) частично новые (новыми	Предложения, сформулированные в диссертационной работе для практического применения, являются полностью новыми, оригинальными и ориентированы на реальные потребности цифровой обработки казахского языка. Они основаны на авторских разработках, специально адаптированных к морфологическим и синтаксическим особенностям казахского языка. Впервые предложен комплексный подход к построению языкового фреймворка QNLP, обеспечивающего интеграцию модулей морфологического анализа, синтаксической и семантической обработки в единую систему. Существенной частью практических рекомендаций является разработка автоматизированных инструментов преданнотации и визуальной разметки текстов, что открывает широкие перспективы для создания масштабируемых

		являются 25-75%); 3) не новые (новыми являются менее25 %)	и высококачественных языковых ресурсов. Представленные решения могут быть эффективно применены при внедрении NLP-технологий в государственные информационные системы, образовательные платформы, интеллектуальные поисковые механизмы и коммерческие цифровые сервисы, ориентированные на казахский язык.
10.	Качество написанияи оформления	Качество академического письма: 1) высокое; 2) среднее; 3) ниже среднего; 4) низкое.	Качество академического письма в диссертации является высоким, текст характеризуется логичной структурой, последовательной подачей материала и точностью формулировок. Изложение ведётся в научном стиле с корректным использованием терминов из областей лингвистики, компьютерных наук и искусственного интеллекта. Все теоретические положения подкреплены аргументацией, а выводы - обоснованы результатами исследования. Чёткая структура и ясность изложения способствуют эффективному раскрытию содержания и демонстрируют высокий уровень академической культуры, обеспечивая научную убедительность и легкость восприятия представленного материала.
11.	Замечания к диссертации		*В литературном обзоре преобладают англоязычные работы, что, с одной стороны, отражает международный уровень исследования, но с другой - недостаточно учитывает локальный контекст. Следовало бы шире привлечь исследования, выполненные на казахском и русском языках в области лингвистики, лексикографии и казахской морфологии, в том числе работы отечественных исследователей. Изложенные выше замечания имеют рекомендательный характер и не оказывают существенного влияния на научную значимость и прикладную ценность полученных результатов.*
12.	Научный уровень статей докторанта по теме исследования (в случае защиты диссертации в форме серии статей официальные рецензенты комментируют научный уровень каждой статьи докторанта по теме		Научный уровень публикаций соискателя полностью соответствует предъявляемым требованиям к диссертационным работам такого уровня

	исследования)		
13.	Решение официального рецензента (согласно пункту 28 настоящего Типового положения)		Ходатайствовать перед Комитетом для присуждения докторанту степени доктора философии (PhD)

Заключение

Представленная диссертация на соискание степени доктора философии (PhD) **Әйтім Әйгерім Қайратқызы** на тему **«Models and methods for the automatic processing of unstructured information»** («Модели и методы автоматической обработки неструктурированной информации»), является законченной самостоятельной научно-исследовательской работой, которая отвечает требованиям «Правил присуждения степеней», а её автор заслуживает ходатайствовать перед Комитетом о присуждении степени доктора философии (PhD).

Официальный рецензент:

PhD, профессор, заместитель генерального директора
Института информационных и вычислительных технологий **Мамырбаев Оркен Жумажанович**

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
«АҚПАРАТТЫҚ ЖӘНЕ ЕСЕПТЕУ ТЕХНОЛОГИЯЛАРЫ ИНСТИТУТЫ» ШАРУАШЫЛЫҚ ЖҮРГІЗУ ҚҰҚЫҒЫНДАҒЫ РЕСПУБЛИКАЛЫҚ МЕМЛЕКЕТТІК КӘСІПОРНЫ
ҚАРЖЫЛЫҚ ЕМЕС ҚҰЖАТТАР ҮШІН

Копии отзывов официальных рецензентов вручаются докторанту не позднее, чем за 5 (пять) рабочих дней до защиты диссертации.