ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

UDC 004.8:004.912
DOI https://doi.org/10.32782/2663-5941/2025.1.2/03

Bautina M.V.
Data Scientist, SoftServe

ADVANCING ENTERPRISE-SCALE INFORMATION RETRIEVAL AND COMPREHENSION USING LARGE LANGUAGE MODELS

The article analyses the implementation of natural language query processing systems for automating information retrieval in enterprises from the perspective of modern information technologies and data science. The study's relevance is determined by the need to optimize access to large volumes of data and enhance the efficiency of management processes by integrating intelligent systems for processing user queries. It has been established that such systems facilitate quick access to relevant information by interpreting natural language queries and generating structured responses. However, problems related to the contextual understanding of queries, multilingual processing, adaptation to domain-specific terminology, and maintaining performance under heavy loads have been identified.

The article aims to analyze the key issues in the operation of natural language query processing systems and develop recommendations for improving their efficiency to optimize information processes in enterprises.

Methodology. *The research is based on a systematic analysis of query processing algorithms, an evaluation of language model performance, and a comparative analysis of accuracy, completeness, and query execution time for queries of varying complexity. Risk analysis methods were used to identify key issues and build optimization models.*

Scientific novelty. *The article comprehensively analyzes the applied aspects of using language models for natural language query processing. The main system limitations have been identified, and measures have been proposed to improve response accuracy and processing speed. Special attention is given to algorithms for automatic query clarification and recommendations for adapting language models to domain-specific data.*

Conclusions. *It has been proven that implementing natural language query processing systems improves the efficiency of information retrieval, provided algorithm transparency and personalization mechanisms are ensured. Recommendations have been made for training models on domain-specific texts, implementing data aggregation algorithms, and using hybrid solutions to optimize query processing in multi-user environments.*

Key words: *artificial intelligence, machine learning, natural language query processing, information retrieval automation, language models, intelligent systems, corporate data, query clarification algorithms, multilingualism, personalization.*

A general statement of the problem and its connection with important scientific or practical tasks. The development of modern artificial intelligence and machine learning technologies opens up new opportunities for optimizing information retrieval and analyzing large amounts of data at enterprises. The increase in the amount of information circulating in corporate systems complicates the process of quick access to relevant data, which affects the efficiency of management decision-making and reduces the competitiveness of enterprises. The main problem is the need to integrate efficient natural language processing models that can not only find relevant information but also provide a deep understanding of it in the context of specific user requests. Big language models show significant potential for improving the accuracy of information retrieval, as they are able to analyze text data with respect to semantic relationships and context. However, the introduction of such technologies at enterprises is accompanied by a number of scientific and practical challenges related to adapting models to industry specifications, increasing their performance, and reducing data processing costs. The issues of ensuring the confidentiality of the information and minimizing the risks associated with the use of automated solutions remain relevant, which requires the development of reliable data protection mechanisms. Thus, the study of the possibilities of using large language models for information retrieval and data understanding in corporate environments is an important step toward the development of innovative solutions for automating processes and increasing the productivity of enterprises.

Analysis of the latest research and publications. Analyzing works devoted to advancing information retrieval and data comprehension at enterprises using large-scale language models demonstrates diverse approaches to optimizing these processes. F. Borges and his team [1] explore approaches to interpreting natural language queries in corporate search engines, which can improve the accuracy and relevance of results. T. Brown and his colleagues [2] examine the effectiveness of few-shot learning for processing queries with a minimum number of training examples, contributing to models' greater versatility.

The BERT model, developed by J. Devlin and co-authors [3], was an important breakthrough due to its bidirectional learning, which significantly improved the understanding of queries' context.

S. Gupta and his colleagues [4] discuss in detail the use of language models to analyse the tone of texts, which opens up new opportunities for monitoring corporate communications. The study by J. Li and co-authors [5] focuses on using deep learning for entity recognition, which helps improve search accuracy.

P. Liu, X. Qiu, and X. Huang [6] propose recurrent neural network algorithms for multitasking text processing, which provide efficiency in processing large amounts of data. C. Montgomery and his team [7] presented an approach to building systems capable of processing natural language queries during peak loads.

C. Raffel and his co-authors [8] investigated the possibilities of transfer learning for generating answers to complex multifactorial queries, which allows for significantly improved results. I.H. Sarker [9] analyzes practical algorithms for search optimization aimed at reducing model bias in detail.

The work of sun c. And his team [10] focuses on tuning bert models for classifying natural language queries in complex databases. Y. Wang and colleagues [11] consider the integration of large language models for medical data transformation, which contributes to the performance of information systems.

The study by a. Wong and co-authors [12] describes methods for automating data retrieval in corporate systems using nlp algorithms. Xlnet, presented by Z. Yang [13], efficiently considers the dependencies between query elements to generate accurate answers. Finally, L. Yu and others [14] highlight the experience of implementing language models to work with large corporate systems using closed databases.

Thus, research shows that large language models can significantly improve search efficiency in corporate environments. However, improving algorithms to ensure high accuracy, speed, and context support for queries is important. Despite significant advances in natural language query processing, the issues of building complex SQL/No-SQL queries with multi-level logical structures remain unresolved, which limits search accuracy and information access efficiency. The adaptation of language models to highly specialized vocabulary and multilingual queries has not been sufficiently studied, which reduces the effectiveness of systems in multilingual corporate environments.

An important challenge remains to ensure the transparency of algorithms and minimize algorithmic bias, which reduces user confidence in automated systems. Existing algorithms for contextual query refinement are not always able to correct errors without losing meaning, which affects the quality of answers. The proposed research aims to address these gaps by developing flexible algorithms for processing complex queries, training models on highly specialized texts, and implementing algorithm auditing methods to increase the transparency of systems. This will improve the accuracy of query processing, support multilingualism, and create an effective information environment to support management decision-making.

The article aims to study the possibilities of using large language models to automate information retrieval at enterprises by building chatbots capable of converting natural language queries into database query language commands, performing data analysis, and providing users with structured answers instead of raw data.

Objectives of the article:

1. To study methods of automating information retrieval at enterprises using large-scale language models and their role in optimizing business processes, including integration with relational and non-relational databases.

2. Describe a chatbot algorithm for transforming natural language queries into SQL/NoSQL queries, processing data, and generating answers, supplemented with a conceptual diagram.

3. Analyse the problems of developing and implementing such systems, including the limitations of language models, query accuracy, and response quality, and offer recommendations for improvement.

Summary of the main material. Modern methods of automating information retrieval at enterprises have been significantly transformed due to the introduction of advanced technologies, particularly large language models. These models, built based on artificial intelligence, can significantly simplify obtaining, analyzing, and presenting data to users. They provide an understanding of natural language queries and transform them into structures suitable for database execution. This opens up new business opportunities where information retrieval is the basis for making effective decisions. Using such systems helps optimize business processes, reduce the time spent on manual data analysis and increase the accuracy of the information received (Table 1).

The application of these methods in modern conditions demonstrates their practical significance. For example, a system based on SQL generation through NLP can be used to automate the creation of sales reports. A manager enters a query such as "Show sales for the last quarter in region X," after which the system generates a corresponding SQL query, retrieves data from the database, structures it, and presents it as a graph or table. At the same time, integration with No-SQL databases is indispensable for companies that work with large amounts of unstructured information, such as analysing customer feedback or system logs.

Practical cases also demonstrate the benefits of chatbots, which, using large language models, provide quick access to data. For example, a chatbot can promptly provide an HR employee with information on the number of open vacancies by transforming a natural language query into a structured response. The real-time analytics provided by such systems allow you to track changes in business processes, which is critical for risk management.

Thus, modern methods of automating information retrieval, with the help of large language models, simplify the process of working with data and increase the efficiency of business processes by adapting information systems to the needs of the modern dynamic environment.

Language models, such as GPT, BERT, and their analogs, have revolutionized natural language processing, providing high accuracy in analyzing and understanding natural language queries. The basic principle of such models is using transformers that allow one to consider the context of words in the query, analyze their relationships, and provide accurate interpretation. This makes them extremely useful for automating processing information requests in a business environment. Integration of language models with databases adds even more capabilities for enterprises, allowing them to combine the simplicity of natural language with the power of querying complex structured and unstructured databases (Table 2).

Modern language models in business practice allow for solving complex information retrieval tasks. For example, GPT can be used to automate processes in financial institutions. A customer enters a query: "What is the account balance for the last month?" The model interprets the query, generates an SQL query to the database, extracts information, and provides a structured answer.

BERT, in turn, can help classify a large number of requests. For example, in the customer support department, it automatically categorises queries into "technical issue," "financial issue 2," or "general query" and then redirects them to the appropriate specialist.

OpenAI Codex is an indispensable tool for developing automated reports. For example, in a logistics company, a Codex-based system automatically generates SQL queries to monitor the status of transportation, significantly reducing the time spent on data analysis.

Table 1

Modern methods of automating information retrieval at enterprises and their application

Method	How it works	Application in practice
Using SQL generation through NLP	The model recognises a natural language query, transforms it into a SQL query, and searches a structured database.	Automation of internal analytical processes, creation of dashboards for management.
Integration with No-SQL databases	The query adapts to No-SQL databases and works with unstructured data (documents, graphs, etc.).	Work with complex data such as reports, system logs, and multi-layered documents.
Using chatbots based on LLM	The user enters a query, the chatbot processes it, extracts data from the database, analyses it, and provides a clear answer.	Customer support, automation of repetitive tasks in commercial organisations.
Integration of real-time analytics	Models work with data streams and analyse them in real time.	Operations monitoring, risk management, real-time business process analytics.

Source: compiled by the author on the basis of [2; 5; 11; 12; 14]

Table 2

Principles of Operation of Language Models and Their Integration with Relational and Non-Relational Databases

Model	How requests are processed	Integration with databases
GPT	It uses an autoencoder to generate text in response to a query, which is capable of understanding the context.	Integrates via API to create SQL queries or analyse text data.
BERT	It uses bidirectional text analysis, taking into account the context on both sides of the query.	It is used to classify queries and build accurate search indexes.
OpenAI Codex	It is focused on generating program code, including SQL queries.	It is used to create complex queries to SQL/No-SQL databases.
Tuned language models	They adapt to the specifics of the domain and analyse requests based on industry specifics.	Ensure high accuracy when working with industry databases.

Source: compiled by the author on the basis of [1; 4; 6; 9; 13]

The practical application of tuned models is especially noticeable in medical institutions where specific terms and contexts require high accuracy. For example, an adapted language model can be used to search for information in large medical research databases based on natural language queries from doctors.

Thus, integrating language models with databases provides an effective mechanism for working with large amounts of information, automating routine tasks, and improving the quality of customer service in various fields.

Chatbots for natural language query processing automate data access by converting user text into structured SQL or No-SQL queries and generating answers in an understandable format. Such systems consist of several key modules that ensure consistent operation: the natural language processing module identifies the essence and intent of the user's query, the query generator forms a SQL or No-SQL query depending on the type of database, the query execution module sends the query to the database to obtain results, and the result processing module converts the received data into a response that is understandable to the user. The algorithm works in stages: After the user enters the text, the system performs its semantic analysis, highlighting key parameters, generates a query to search for data, and returns a structured response after execution. In practical terms, such a system provides quick access to data, automatically processing even complex queries. Figure 1 shows the sequence of interaction between the main components of the algorithm, illustrating the key stages of information processing and data transfer between modules.

The natural language query processing algorithm starts with the user entering text. The natural language processing module analyses the text to identify key entities and query intentions. The identified parameters are passed to the query generator, which generates a structured SQL or No-SQL query depending on the specifics of the database. The query is sent to the database, where the information is searched, and the results are sent to the processing module. This module converts the data into a user-friendly format and returns the answer, for example, as a table, text message, or graphical display.

To test the chatbot's efficiency, a research experiment was conducted to assess the accuracy and speed of processing natural language queries when interacting with a database of scientific publications. The experiment aimed to determine whether the system could correctly form SQL and No-SQL queries and provide answers in a structured and clear format. The experiment was conducted under standard conditions of access to a database with real-time scientific articles and reports. Three types of queries were tested: simple queries with one parameter, queries of medium complexity with several conditions, and complex queries with typos and incomplete phrases.

Simple queries included searching for data by year of publication or the author's name. Medium-complexity queries contained several criteria, such as "show all articles on neural networks in the last three years." Complex queries with errors involved entering incomplete queries or words with typographical errors, such as "publications instead of publication." For each query type, the processing time, response accuracy, and data completeness were recorded to assess how the algorithm responds to both correct and incorrect wording.

Table 3 presents the results of testing the effectiveness of the natural language query processing algorithm.

The accuracy metric was calculated as the percentage of correctly generated responses divided by the total number of requests for a specific query type. The formula used for this calculation was:

$$\text{Accuracy}(\%) = \frac{(\text{Number of correct responses})}{(\text{Total requests})} * 100\% \quad (1)$$

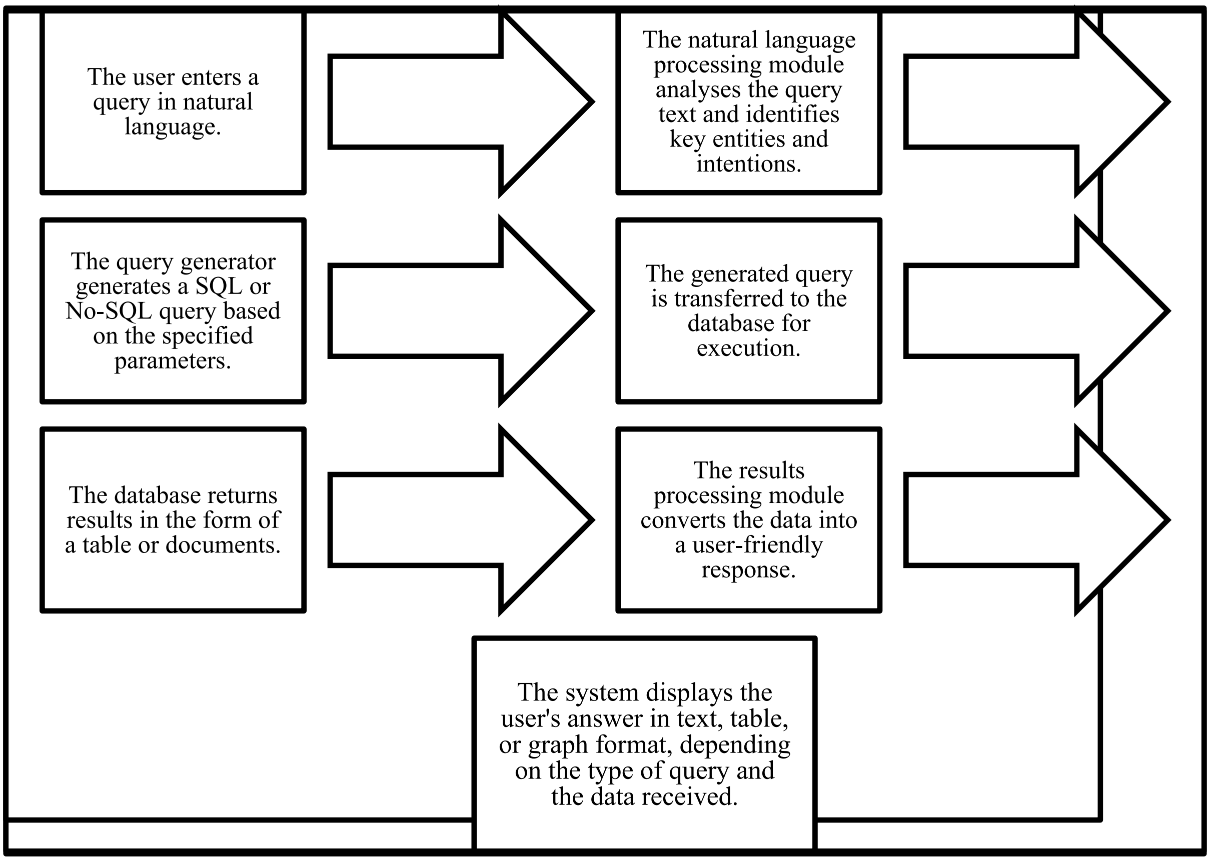

Fig. 1. The sequence of the query transformation algorithm

Source: author's development

Table 3

Results of testing the effectiveness of the natural language query processing algorithm

Request type	Number of requests	Accuracy, %	Completeness, %	Processing time, s
Simple query with one parameter	20	97	96	0,7
Medium complexity request	20	93	90	1,1
Request with errors or incomplete wording	10	82	78	1,6

Source: author's development

For example, if 19 out of 20 simple queries returned correct results, the accuracy would be:

$$\text{Accuracy} = \frac{19}{20} * 100\% = 95\%$$

Completeness measured the proportion of retrieved data relative to the total expected data. It was calculated as follows:

$$\text{Completeness}(\%) = \frac{\text{Amount of retrieved data}}{\text{Total expected data}} * 100\% \quad (2)$$

For instance, if a medium-complexity query retrieved 45 out of 50 expected records, the completeness would be:

$$\text{Completeness}(\%) = \frac{45}{50} * 100\% = 90\%$$

Processing time, reflecting the average time taken to process each query type, was calculated as:

$$\text{Processing Time}(s) = \frac{\text{Total processing time for all requests}}{\text{Number of request}} \quad (3)$$

The experiment results showed that the natural language query processing system algorithm demonstrates high accuracy when executing simple queries with clearly defined parameters. However, when processing complex queries containing numerous conditions, logical operators, and filters, a decrease in accuracy, increased processing time, and a complication in generating a relevant answer was observed. These results indicate several problems related to the functioning of language models, the optimization of computing processes, and the system's ability to interpret contextual relationships between query elements.

One of the main problems is the limitation of language models used for query analysis. Most modern models are trained on large publicly available corpora of texts, which often do not take into account the specifics of individual fields of knowledge and professional terminology [3]. This leads to difficulties in understanding highly specialized vocabulary and complex syntactic constructions. It is especially problematic to process ambiguous words and terms with different meanings depending on the context. In scientific and corporate environments, this problem manifests when forming queries containing specific abbreviations, scientific terms, or multi-component constructions. As a result, the system can incorrectly identify query entities and incorrectly form SQL or No-SQL queries, significantly reducing the answer's accuracy.

The problem of contextual understanding is also a key factor affecting system performance. Natural language processing algorithms try to identify key entities and parameters. However, when processing polysemous words or queries with multiple logical conditions, the system may lose the connection between individual query elements [6]. This leads to errors in query generation and irrelevant results. Such situations are typical for queries with time frames, numeric values, and comparison operators that require processing complex logical dependencies. For example, a query such as "Show all project progress reports for the last three years that have been approved but do not contain budget changes" may be misinterpreted if the system cannot combine all parameters in the correct order.

Another important issue is the quality of the answer that the system returns to the user. In many cases, the result of query processing is presented in the form of a large amount of data, which makes it difficult to perceive information and inconvenient to search for key results. For example, a query about financial indicators for a certain period may return a large amount of data without aggregation, forcing the user to analyze each record independently [9]. This reduces system efficiency and increases decision-making time. Thus, it is necessary to implement mechanisms that allow filtering and aggregating data so that the response contains only the key information relevant to the user's request.

Computing resource limitations also significantly affect system performance, especially when processing many simultaneous requests. In systems with large amounts of data, query processing time can increase during peak load, which creates delays in generating a response. This is critical for businesses where timely access to data is a key factor in successful operations, such as financial services and monitoring systems. Delays can slow down decision-making processes and increase the risk of errors due to untimely data updates.

Another problem is the system's resistance to incorrect or incomplete queries. Users may make typos, enter incomplete wording, or use non-standard language constructions. The absence of contextual clarification algorithms means the system may return false results or provide no answer. This reduces user confidence in the system and may limit its use in critical tasks. Therefore, it is necessary to integrate algorithms that can automatically correct typos, offer refined query options, or generate queries based on the analysis of previous requests.

Support for multilingualism also remains an important task for system improvement. Most language models are optimised for the English language, which limits the system's effectiveness in multilingual environments [11]. To process queries in other languages, additional training of models on relevant corpora of texts that take into account language and cultural peculiarities is required. This will ensure the accuracy of query processing in different languages and avoid errors in recognizing grammatical structures or words with multiple meanings.

To improve chatbot performance in the context of automating information retrieval, it is recommended that comprehensive measures be implemented, including adapting language models to specific industry needs by training on corporate texts and documents. This will allow the system to process highly specialized queries more efficiently and improve the accuracy of identifying key entities. An important step is introducing refinement and contextual analysis algorithms that can correct typos and generate query refinement options. It is advisable to use hybrid solutions that combine relational databases with caching and distributed data storage systems to accelerate query processing. This will optimize system performance even in high load conditions. Filtering mechanisms and automatic aggregation of results will help create clear and concise answers that provide users access to key information without overloading them with unnecessary data.

Personalized chatbot algorithms allow you to adapt their functions to individual user needs, considering the history of requests and generating suggestions based on typical search scenarios. This improves the system's usability and speeds up access to the necessary data. The prospects for the use of such chatbots cover various industries: in the financial sector, they speed up the creation of analytical reports and simplify access to key indicators; in the logistics sector, they facilitate the tracking and management of operations; in educational institutions, they provide quick access to educational materials and research results. In the long term, such systems will contribute to creating an automated information environment that ensures efficient interaction with large amounts of data and optimization of internal processes at enterprises.

Conclusions. It has been established that using language models in natural language query processing systems increases the efficiency of information retrieval automation but is accompanied by several limitations. The main problems are insufficient contextual understanding of complex queries, limited adaptation to highly specialized vocabulary, difficulties with multilingualism, and increased processing time under high load. The system can also generate large amounts of irrelevant data, which makes it difficult for the user to get a clear answer. To address these shortcomings, we propose additional training of models on industry texts, algorithms for contextual query refinement, and hybrid caching and filtering systems to optimize data processing. Introducing personalised algorithms will allow the history of queries to be considered to generate more accurate answers.

Further research should focus on providing multilingual support, developing mechanisms for algorithmic transparency and user confidence, and improving the system to adapt answers to the query context. These measures will create an effective information environment for making management decisions based on accurate data.

Bibliography:

1. Query Understanding for Natural Language Enterprise Search / F. Borges et al. *arXiv preprint*. 2020. Available at: https://arxiv.org/abs/2012.06238 (date of access: 10.01.2025).

2. Language Models are Few-Shot Learners / T. Brown et al. *Advances in Neural Information Processing Systems*. 2020. Vol. 33. P. 1877-1901. Available at: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html (date of access: 10.01.2025).

3. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*. 2018. Available at: https://arxiv.org/abs/1810.04805 (date of access: 10.01.2025).

4. Gupta S., Singh A., Kumar V. Emoji, Text, and Sentiment Polarity Detection Using Natural Language Processing. *Information*. 2023. Vol. 14, no. 4. P. 222. URL: https://doi.org/10.3390/info14040222 (date of access: 12.01.2025).

5. A Survey on Deep Learning for Named Entity Recognition / J. Li et al. *IEEE Transactions on Knowledge and Data Engineering*. 2020. P. 1. URL: https://doi.org/10.1109/tkde.2020.2981314 (date of access: 12.01.2025).

6. Liu P., Qiu X., Huang X. Recurrent Neural Network for Text Classification with Multi-Task Learning. *arXiv preprint*. 2016. Available at: https://arxiv.org/abs/1605.05101 (date of access: 10.01.2025).

7. Montgomery C., Isah H., Zulkernine F. Towards a Natural Language Query Processing System. *2020 1st International Conference on Big Data Analytics and Practices (IBDAP)*, Bangkok, Thailand, 25–26 September 2020. 2020. URL: https://doi.org/10.1109/ibdap50342.2020.9245462 (date of access: 12.01.2025).

8. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer / C. Raffel et al. *Journal of Machine Learning Research*. 2020. Vol. 21, No. 140. P. 1-67. Available at: https://www.jmlr.org/papers/v21/20-074.html (date of access: 10.01.2025).

9. Sarker I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2021. Vol. 2, no. 3. URL: https://doi.org/10.1007/s42979-021-00592-x (date of access: 12.01.2025).

10. How to Fine-Tune BERT for Text Classification? / C. Sun et al. *Lecture Notes in Computer Science*. Cham, 2019. P. 194–206. URL: https://doi.org/10.1007/978-3-030-32381-3_16 (date of access: 12.01.2025).

11. An integrated big data analytics-enabled transformation model: Application to health care / Y. Wang et al. *Information & Management*. 2018. Vol. 55, no. 1. P. 64–79. URL: https://doi.org/10.1016/j.im.2017.04.001 (date of access: 12.01.2025).

12. A Survey of Natural Language Processing Implementation for Data Query Systems / A. Wong et al. *2021 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, Shanghai, China, 12–14 December 2021. 2021. URL: https://doi.org/10.1109/rasse53195.2021.9686815 (date of access: 12.01.2025).

13. Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R., Le Q.V. XLNet: Generalised Autoregressive Pretraining for Language Understanding. *arXiv preprint*. 2019. Available at: https://arxiv.org/abs/1906.08237 (date of access: 10.01.2025).

14. Experience with Large Language Model Applications for Information Retrieval from Enterprise Proprietary Data / L. Yu et al. *Lecture Notes in Computer Science*. Cham, 2024. P. 92–107. URL: https://doi.org/10.1007/978-3-031-78386-9_7 (date of access: 12.01.2025).

Баутіна М.В. АВАНСУВАННЯ ІНФОРМАЦІЙНОГО ПОШУКУ ТА ОСМИСЛЕННЯ ДАНИХ НА ПІДПРИЄМСТВАХ ЗА ДОПОМОГОЮ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

У статті проаналізовано впровадження систем обробки природномовних запитів для автоматизації інформаційного пошуку на підприємствах з позиції сучасних інформаційних технологій та науки про дані. Актуальність дослідження обумовлена необхідністю оптимізації доступу до великих обсягів даних і підвищення ефективності управлінських процесів шляхом впровадження інтелектуальних систем для обробки запитів користувачів. Встановлено, що такі системи сприяють швидкому доступу до релевантної інформації завдяки інтерпретації природномовних запитів і формуванню структурованих відповідей. Разом із тим, виявлено проблеми, пов'язані з контекстуальним розумінням запитів, обробкою багатомовних запитів, адаптацією до вузькоспеціалізованої лексики та забезпеченням продуктивності при високих навантаженнях.

Метою статті є аналіз ключових проблем роботи систем обробки природномовних запитів та розробка рекомендацій щодо підвищення їхньої ефективності для оптимізації інформаційних процесів на підприємствах.

Методологія. Дослідження проведено на основі системного аналізу алгоритмів обробки запитів, оцінювання ефективності роботи мовних моделей та порівняльного аналізу показників точності, повноти та часу виконання запитів різного рівня складності. Використано методи аналізу ризиків для виявлення ключових проблем і побудови моделей оптимізації.

Наукова новизна. У статті представлено комплексний аналіз прикладних аспектів використання мовних моделей для обробки природномовних запитів. Виявлено основні обмеження системи та запропоновано заходи для підвищення точності відповідей і швидкості обробки. Окрему увагу приділено алгоритмам автоматичного уточнення запитів та рекомендаціям щодо адаптації мовних моделей до галузевих даних.

Висновки. Доведено, що впровадження систем обробки природномовних запитів сприяє підвищенню ефективності інформаційного пошуку за умови забезпечення прозорості алгоритмів та впровадження механізмів персоналізації. Запропоновано рекомендації щодо навчання моделей на вузькоспеціалізованих текстах, впровадження алгоритмів автоматичної агрегації даних та використання гібридних рішень для оптимізації обробки запитів у багатокористувацьких середовищах.

Ключові слова: *штучний інтелект, машинне навчання, обробка природномовних запитів, автоматизація інформаційного пошуку, мовні моделі, інтелектуальні системи, корпоративні дані, алгоритми уточнення, багатомовність, персоналізація.*