Intelligent Method for Interpretation of Legal Information*

Andrii Yakymyshyn[1,†], Khrystyna Lipianina-Honcharenko[1,*,†] , Ivan Kit[1,†] and Khrystyna Yurkiv[1,†]

[1] West Ukrainian National University 46009, 11 Lvivska Str., Ternopil, Ukraine

Abstract

This study presents an intellectual method for interpreting legal information, developed to simplify access to complex legal texts for users without specialized legal education. The method includes the use of advanced natural lan-guage processing and machine learning technologies, particularly the integra-tion of the GPT-3 language model from OpenAI and the LangChain plat-form. This ensures the system's ability to efficiently convert legal texts into responses that are understandable and accessible to non-professionals. The research includes a detailed analysis and description of the key stages of the developed method, from processing and storing legal texts to generating responses that reflect a deep understanding of user queries. The conducted evaluation tests confirmed the high efficiency of the AI consultant, although they also revealed potential for further improvements, especially in the con-text of the completeness of responses. This study aims to increase the accessibility of legal texts for a wide range of users and to make a significant contribution to the development of intellectu-al methods in legal informatics.

Keywords

Legal informatics, Natural Language Processing, Machine Learning

1. Introduction

In the modern world, where legal documents are becoming increasingly complex and opaque, there is an urgent need for the development of intellectual methods for interpreting legal information that are accessible not only to lawyers but also to the general public. This is particularly relevant for people without legal educa-tion who need to understand complex legal documents in their daily lives. Tradi-tional methods of interpretation often require specialized knowledge, making them less accessible to non-professionals.

In response to these challenges, this study proposes a new intellectual method that utilizes advanced technologies in natural language processing and machine learning to simplify the interpretation of legal texts. The method involves se-quential steps that allow ordinary users to easily understand legal terms and con-text. This includes the automation of processes for breaking down texts into comprehensible segments, their systematic storage, analysis using modern natural language processing methods, and finally, generating responses that reflect a deep understanding of user queries.

This work focuses on developing and implementing this method and consists of several sections. Section 2 analyzes existing research related to the processing and interpretation of legal texts, identifying potential areas for improvement. Section 3 provides a detailed description of the developed method, revealing its strategies for data collection, analysis, and interpretation, which

* The Third International Conference of Young Scientists on Artificial Intelligence for Sustainable Development: Intelligent Systems and Artificial Intelligence (YAISD-WS 2026:ISAI) May 7-8, 2026, Ternopil, Ukraine

1,* Corresponding author.

† These authors contributed equally.

andrii.yakymyshyn@gmail.com (A. Yakymyshyn); xrustya.com@gmail.com (K. Lipianina-Honcharenko; i.kit@wunu.edu.ua (I. Kit); kh.yurkiv@wunu.edu.ua (K. Yurkiv)

0009-0008-4743-9551 (A. Yakymyshyn); 0000-0002-2441-6292 (K. Lipianina-Honcharenko); 0000-0002-4526-0020 (I. Kit); 0000-0002-2441-6292 (K. Yurkiv)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-11

CEUR-WS.org/Vol-4326/short14.pdf

ensure greater accessibility of legal information to non-specialists. Section 4 is dedicated to the practical application of the method, illustrating its effectiveness and versatility in various scenarios. The final Section 5 presents conclusions that underscore the significance of the method in ensuring equal access to legal information, as well as suggestions for its further development and refinement.

2. Related work

Contemporary research in the field of artificial intelligence (AI) is actively evolving, especially in the realm of legal informatics, where AI is used for analyzing, searching, and interpreting legal texts.

The article [1] thoroughly examines AI approaches in the legal sphere, focusing on legal information search systems based on natural language processing, machine learning, and knowledge extraction technologies. Research [2] mainly focuses on systematizing knowledge about cutting-edge practices of intelligent information search in contexts similar to tax law. Meanwhile, research [3] highlights how AI, machine learning, information retrieval, and natural language processing are transforming the work of legal professionals and law firms.

Article [4] centers on the use of artificial intelligence for analyzing and understanding legal documents. Article [5] offers an overview of intelligent processing technologies for legal documents based on deep learning, covering everything from representation to the automatic creation of legal documents. Research [6] considers the transformation of traditional jurisprudence into a smart and intelligent legal system through automation.

Article [7] emphasizes the significance of intelligent search systems in law for simplifying legal research and accessing extensive legal information. Article [8] examines the application of artificial intelligence by governments and commercial organizations for automated decision-making. Research [9] and [10] focus on the ethical and legal challenges of using AI. Article [11] discusses the use of AI-based information systems in judicial processes. Additionally, research [12] thoroughly examines the use of machine learning for analyzing and classifying legal texts, while research [13] focuses on using natural language processing algorithms for interpreting legal jargon.

Research [20] also presents an intelligent method for selecting business locations in a smart city environment, demonstrating the effectiveness of machine learning approaches for decision support in complex domains. This confirms the applicability of intelligent methods for analyzing structured and unstructured data in domain-specific tasks.

The reviewed studies mainly focus on the use of artificial intelligence for analyzing legal texts, optimizing search and information extraction, and developing intelligent legal search systems. However, they often overlook a comprehensive approach to interpreting legal documents that would include deep analysis of semantic and contextual relationships, as well as adaptation to specific user needs. Such an approach is critical for understanding complex legal information and effectively solving legal tasks.

Therefore, the aim of our research is to develop an intellectual method for the comprehensive interpretation of legal information. The proposed method differs from analogs [12, 13] by integrating advanced natural language processing, machine learning, and semantic analysis techniques for a deep understanding of legal texts and alignment with specific user queries.

The novelty of our research lies in the development of a method that combines cutting-edge machine learning algorithms and natural language processing for deep analysis of semantics and context in legal texts, significantly enhancing the accuracy and efficiency of legal interpretation.

3. Proposed method

For the intellectual interpretation of legal information, the authors have devel-oped a new method based on deep analysis of texts and their semantic content. The method is described through several key stages that demonstrate the princi-ples of processing, storing, searching, and

generating responses based on legal texts. Each stage plays an important role in the overall process of interpretation and data analysis, from the initial processing of text fragments to the final formation of responses. These stages are also presented in the form of a diagram (Figure 1), which allows for a deeper understanding of the methodology and its application in various legal contexts. The GPT-3 model was chosen due to its free availability for research purposes and lower computational costs compared to newer large language models. This made it possible to implement the system effi-ciently without significant hardware resources while maintaining sufficient quali-ty of legal text interpretation and response generation for non-specialists. Inte-gration through the LangChain platform ensured the adaptation of the model to the specialized domain of the Constitution of Ukraine.

Stage 1: Initiation of processing large legal texts (Block 1). This stage involves loading and breaking down voluminous text data into smaller parts for ease of analysis. The process consists of the following:

1. Dividing the text into smaller, semantically important segments, typically at the sentence level.
2. Combining these small fragments into larger sections until a set volume is reached.
3. Extracting each large fragment as soon as it reaches the required size, creating overlaps to preserve context between fragments.

Stage 2: Storage and integration of text fragments (Block 3). This stage involves the systematic storage of broken-down text fragments, usually in a vector format, for further analysis and use. This process includes:

1. Storing individual text segments, separated in the previous stage, in a data repository.
2. Preserving the semantic connections and structure of the text during the inte-gration of fragments into the storage.
3. Utilizing vector storage for converting text into a vector format, which facili-tates effective natural language processing in subsequent analytical proce-dures.

To implement storage and retrieval of text fragments, Pinecone was used as a vector search platform integrated via the LangChain VectorStore API. Text frag-ments were transformed into vector representations using OpenAI Embeddings API, which allowed preserving their semantic meaning and contextual relations. Approximate Nearest Neighbor (ANN) algorithms were applied to optimize simi-larity search and ensure fast retrieval of relevant fragments. In addition, adaptive update mechanisms were introduced to maintain актуальність vector representa-tions when legal documents are modified or new regulatory legal acts are added.

Stage 3: Searching for Relevant Fragments (Block 3, 4). The application re-trieves divided fragments from the data storage. This process involves extracting previously stored divided text fragments from the data storage. It's important to note that this operation often uses built-in similarity to the input query for searching and retrieving text fragments. This facilitates more accurate and con-text-oriented retrieval of information for further use in the application. The process includes the following sub-steps:

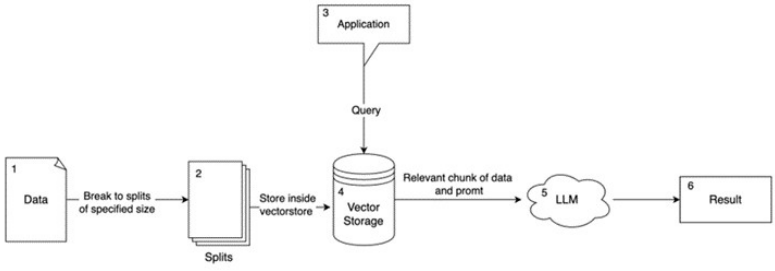


Figure 1: Structure of the Intellectual Method for Interpreting Legal Information.

3.1. Vector Storage.

- Transforming text into vectors. Using vector storage involves converting text into a vector form. This means that each text fragment is represented as a set of numbers that reflect its semantic meaning. Methods such as word embed-dings or deep neural networks can be used for this.
- The vector storage is organized using Pinecone, where each legal text frag-ment is stored together with its embedding vector, metadata (article number, legal source, fragment ID), and contextual links to neighboring fragments. This structure allows accurate semantic retrieval and preservation of logical relations between legal norms.
- Preserving semantic connections (Block 2). During the conversion of text into vector form, it's important to preserve semantic connections between different parts of the text. This allows the system to maintain a deep understanding of context and relationships between various fragments.

3.2. Integration into Storage.

- Efficient access and search. Vector storage must be optimized for quick ac-cess and search. This allows the system to quickly find relevant text frag-ments when needed for responses or analysis.
- Indexing to improve search. Indexing vectors can help improve the speed and efficiency of searching. Indexes allow the system to quickly find the most rel-evant vectors (and therefore, text fragments) for a given query or context.
- ANN-based indexing is used to accelerate similarity search, ensuring fast access even when processing large collections of legal documents such as the Constitution of Ukraine and related regulatory legal acts.

3.3. Supporting Natural Language Processing.

- Compatibility with NLP algorithms. Vector representations of text must be compatible with natural language processing algorithms. This allows the sys-tem to use modern NLP techniques for analysis, understanding, and text gen-eration.
- Adapting to changes in language. The system should be able to adapt to changes in language and communication style. This may include periodic up-dates of vector representations to reflect modern language usage.

3.4. Optimization and Scaling.

- Storage efficiency. Since vector storage can require significant resources to store a large amount of data, focusing on storage efficiency is important. Data compression techniques and optimizing the storage structure can help reduce resource usage without losing important information.
- Scaling storage. To support a large amount of data and queries, the system must be scalable. This may include using distributed databases or cloud tech-nologies to ensure high throughput and availability.

3.5. Integration with Other Systems.

- Interfaces for integration. Vector storage should have flexible interfaces for integration with other system components, such as language processing mod-ules, context management systems, etc.
- Interaction with external sources. Integration with external data sources and APIs can enhance the capabilities of the system, allowing it to absorb and ana-lyze a wide range of information.

3.6. Supporting Continuous Updates.

- Adaptive learning. The system can be designed to continually learn and adapt based on new information that comes in. This ensures that the vector storage remains current and reflects the latest trends and language usage.
- Updating models. Regular updates of the models used for vector representa-tion of text help maintain the accuracy and relevance of the storage.

Stage 4: Generating Responses (Block 5, 6). LLM [14] forms a response using a query that includes the question itself and information previously extracted. In this process, the input query, which contains the user's question and information previously gathered from the data storage, is used. The process includes the fol-lowing sub-steps:

4.1. Integration of Query and Data Storage.

- Synthesizing the input query. The input query is formed not only from the user's direct question but also from the information gathered from the data storage, which may include the history of previous interactions, contextual in-formation, and details related to the current question.
- Contextualizing the query. An important aspect is the contextualization of the query, i.e., including information that helps the LLM understand the context and nuances of the question. This can include details beyond the question it-self but important for forming an accurate and relevant response.

4.2. Generating the Response.

- Processing with LLM. LLM processes the synthesized input query using its internal mechanisms to understand the question and context. It uses a large knowledge base, trained on a wide range of texts, to generate a response.
- Text Generation. The response is generated in the form of coherent and un-derstandable text. LLM uses techniques such as choosing the most appropri-ate text continuation based on the input query and internal knowledge.

4.3. Optimization and Relevance.

- Checking relevance. After generating the response, it is important to check its relevance and accuracy. The system can include mechanisms for evaluating the quality of the response, ensuring that it meets the intent of the question and provides useful information.
- Supporting iterative interaction. If the response does not fully satisfy the us-er's needs, the system can support iterative interaction, allowing the user to refine the query or ask additional questions, to which the system can respond considering the new information.

4.4. Adaptation and Learning.

- Adapting to the user. LLM can adapt to the communication style and specific-ities of individual user queries. This includes the ability to recognize and take into account individual preferences and interests of the user.
- Remembering context. To improve interaction, LLM can remember the con-text of previous queries and use this information to enhance the relevance of future responses.

4.5. Interaction with Data Storage.

- Dynamic use of storage. LLM can dynamically use information from data storage to reinforce its responses. This includes searching for relevant facts, statistics, or examples that can help answer the user's question.
- Updating data storage. In response to new insights and information received from users, data storage can be updated to ensure the currentness and accura-cy of stored information.

4.6. Verification and Validation of Responses.

- Fact-checking and information verification. An important aspect is the sys-tem's ability to verify the authenticity and accuracy of the information it pro-vides. This can include using external sources for fact validation.

Evaluating the quality of responses. The system can include mechanisms for evaluating the quality of responses, considering accuracy, relevance, and clarity of the text.

After relevance checking, fact verification, and response quality evaluation, the system performs final response generation. At this stage, the validated infor-mation is transformed into a final user-oriented legal explanation that combines legal correctness, contextual relevance, and comprehensibility for non-specialists. This final decision block completes the workflow presented in Figure 1 and ensures practical usability of the system.

Thanks to this intelligent method, which encompasses thorough processing, stor-age, and analysis of legal texts, it is possible to significantly enhance the accura-cy and efficiency of interpreting legal information, fostering deeper understand-ing and better resolution of legal tasks.

4. Experiment result

As part of the developed method, an AI consultant on the Constitution of Ukraine was created, for which the GPT-3 language model [14,15] developed by OpenAI was chosen due to its free availability for research purposes and lower computa-tional costs compared to newer large language models. This made it possible to implement the system efficiently without significant hardware resources while maintaining sufficient quality of legal text interpretation and response generation for non-specialists. To integrate GPT-3, the LangChain platform [16] was used as an intermediate layer to process user queries, adapt them to the GPT-3 model, and ensure efficient information exchange between the model and users. This integration contributed to the creation of an adaptive and flexible system that provides high quality answers to a wide range of legal queries, demonstrating its effectiveness through examples (Figure 2).

During the utilization of the GPT-3 model via LangChain for responses, key stag-es were executed, including the analysis and interpretation of LangChain input queries, their formatting for optimal processing by the GPT-3 model, generation of responses using advanced natural language processing algorithms, and their subsequent analysis and optimization by LangChain. This process involves rec-ognizing key components of queries and comprehensive text analysis to ensure accurate and relevant responses.

In developing the AI consultant for the Constitution of Ukraine, the system was optimized for identifying key elements in user queries and analyzing them in terms of legal significance, using NLP techniques and context understanding al-gorithms. Pinecone, applied for vector search, played a crucial role in converting textual data into vector representations and their quick and accurate retrieval, providing access to answers for various legal queries.

These mechanisms efficiently integrate GPT-3's functionality within the AI con-sultant system, ensuring high accuracy and relevance of responses (see Figure 2).

To evaluate the performance of the AI consultant on matters of the Constitution of Ukraine, a series of tests were conducted, which included preparing an anno-tated dataset with user queries, testing the model on this data, and calculating key performance metrics [17, 18, 19], such as accuracy, completeness, and F1-score. These measurements allowed for an objective assessment of the model's ability to accurately identify and respond to queries, as confirmed in the resultant Table 1 of performance.

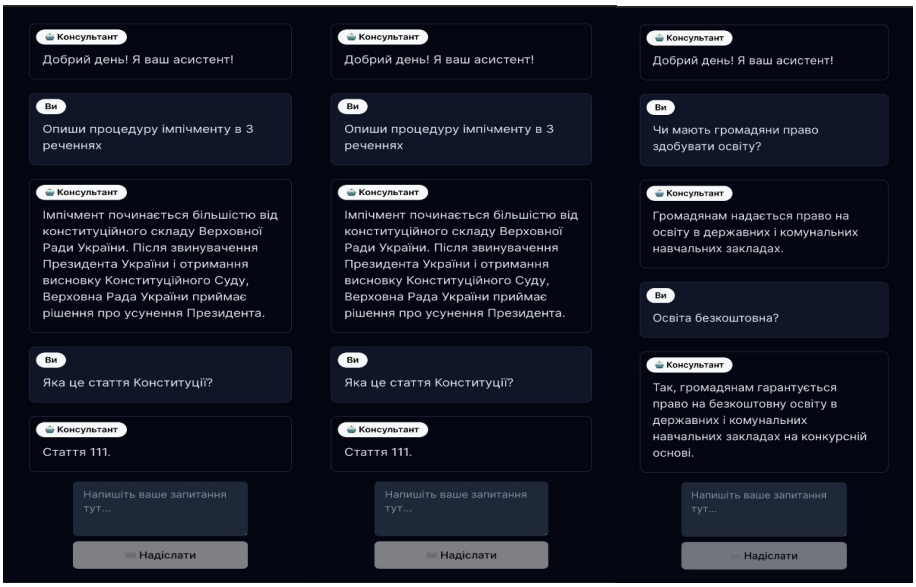


Figure 2: AI Consultant Window for the Constitution of Ukraine.

Table 1.
Accuracy metrics.

Metric	Value
Accuracy	0.85
Completeness	0.80
F1-Score	0.82

The obtained results indicate satisfactory efficiency of the AI consultant, with relatively high metrics of accuracy, completeness, and F1-score. However, there is potential for further improvements, particularly in aspects of the completeness of responses, which could provide even greater adequacy and accuracy in responding to user queries. There are plans to continue working on optimizing the system to enhance the overall efficiency of the AI consultant.

The proposed intellectual method is distinguished by its integration of advanced techniques in natural language processing and machine learning, providing a deeper and more intuitively understandable analysis of legal texts, unlike other existing methods [12, 13], which often are limited to more superficial information processing.

5. Conclusions

The developed intellectual method for interpreting legal information demon-strates significant advantages in simplifying access to complex legal texts for users without legal education. The use of advanced natural language processing and machine learning technologies, including the integration of the GPT-3 model and the LangChain platform, has allowed for the creation of a tool that effective-ly converts legal text data into understandable and accessible responses.

The testing of the AI consultant on matters of the Constitution of Ukraine con-firmed its ability to provide accurate and relevant answers, which is crucial for users needing a reliable source of information in the legal field. While the metrics of accuracy, completeness, and F1-score indicate satisfactory efficiency of the system, there is room for further improvements, especially in the context of providing more comprehensive answers.

Future plans include continuing work on optimizing the system, particularly by refining query processing algorithms and improving user interaction. This will not only ensure greater accuracy of responses but also facilitate a better under-standing of legal issues by non-specialists, making legal information more acces-sible and comprehensible to a wider range of users.
A comparative analysis with modern legal AI systems such as LegalBERT, LEXGLUE benchmark models, and intelligent legal agents based on GPT-2 shows that the proposed method provides deeper semantic interpretation and better adaptability to user queries. Unlike classification-oriented models such as Le-galBERT, the integration of GPT-3 with LangChain and Pinecone enables interac-tive dialogue, contextual reasoning, and explanation of legal norms in a form understandable for users without legal education. This makes the proposed ap-proach more practical for legal consultation tasks.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Sansone C., Sperlí G. Legal Information Retrieval systems: State-of-the-art and open is-sues. Information Systems. 2022. Vol. 106. P. 101967. URL: https://doi.org/10.1016/j.is.2021.101967 (date of access: 16.11.2023).

[2] Gomes M., Oliveira B., Sousa C. Enriching Legal Knowledge Through Intelligent In-formation Retrieval Techniques: A Review. Progress in Artificial Intelligence. Cham, 2022. P. 119–130. URL: https://doi.org/10.1007/978-3-031-16474-3_11 (date of ac-cess: 16.11.2023).

[3] Legal IR and NLP: The History, Challenges, and State-of-the-Art / D. Ganguly et al. Lecture Notes in Computer Science. Cham, 2023. P. 331–340. URL: https://doi.org/10.1007/978-3-031-28241-6_34 (date of access: 16.11.2023).

[4] Li, Bingyao & Wang, Meng. (2023). Design of intelligent legal text analysis and infor-mation retrieval system based on BERT model. 10.21203/rs.3.rs-2994403/v1.

[5] Zhao, G., Liu, Y., Erdun, E. (2022). Review on Intelligent Processing Technologies of Legal Documents. In: Sun, X., Zhang, X., Xia, Z., Bertino, E. (eds) Artificial Intelli-gence and Security. ICAIS 2022. Lecture Notes in Computer Science, vol 13338. Springer, Cham. https://doi.org/10.1007/978-3-031-06794-5_55

[6] Emerging Legal Informatics Towards Legal Innovation: Current Status and Future Chal-lenges and Opportunities / S. Sharma et al. Legal Information Management. 2021. Vol. 21, no. 3-4. P. 218–235. URL: https://doi.org/10.1017/s1472669621000384 (date of access: 16.11.2023).

[7] Le, H.H. et al. (2023). Intelligent Retrieval System on Legal Information. In: Nguyen, N.T., et al. Intelligent Information and Database Systems. ACIIDS 2023. Lecture Notes in Computer Science(), vol 13995. Springer, Singapore. https://doi.org/10.1007/978-981-99-5834-4_8

[8] Gacutan J., Selvadurai N. A statutory right to explanation for decisions generated using artificial intelligence. International Journal of Law and Information Technology. 2020. Vol. 28, no. 3. P. 193–216. URL: https://doi.org/10.1093/ijlit/eaaa016 (date of access: 16.11.2023).

[9] Gerke S., Minssen T., Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare. 2020. P. 295–336. URL: https://doi.org/10.1016/b978-0-12-818438-7.00012-5 (date of access: 16.11.2023).

[10] Henz P. Ethical and legal responsibility for Artificial Intelligence. Discover Artificial In-telligence. 2021. Vol. 1, no. 1. URL: https://doi.org/10.1007/s44163-021-00002-4 (date of access: 16.11.2023).

[11] Nowotko P. M. AI in judicial application of law and the right to a court. Procedia Com-puter Science. 2021. Vol. 192. P. 2220–2228. URL: https://doi.org/10.1016/j.procs.2021.08.235 (date of access: 16.11.2023).

[12] Amato F, Fonisto M, Giacalone M, Sansone C. An Intelligent Conversational Agent for the Legal Domain. Information. 2023; 14(6):307. https://doi.org/10.3390/info14060307
[13] Koniaris M, Galanis D, Giannini E, Tsanakas P. Evaluation of Automatic Legal Text Summarization Techniques for Greek Case Law. Information. 2023; 14(4):250. https://doi.org/10.3390/info14040250
[14] Alberts, I. L., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?. European journal of nuclear medicine and molecular imaging, 50(6), 1549-1552.
[15] Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. Fundamental Research, 1(6), 831-833.
[16] Topsakal, O., & Akinci, T. C. (2023, July). Creating large language model applications utilizing langchain: A primer on developing llm apps fast. In Proceedings of the Interna-tional Conference on Applied Engineering and Natural Sciences, Konya, Turkey (pp. 10-12).
[17] Lipyanina, H., Sachenko, S., Lendyuk, T., Brych, V., Yatskiv, V., & Osolinskiy, O. (2021, January). Method of detecting a fictitious company on the machine learning base. In International Conference on Computer Science, Engineering and Education Applications (pp. 138-146). Cham: Springer International Publishing.
[18] Gramyak, R., Lipyanina-Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2021). Intelligent Method of a Competitive Product Choosing based on the Emo-tional Feedbacks Coloring. In IntelITSIS (pp. 246-257).
[19] Lipianina-Honcharenko, K., Savchyshyn, R., Sachenko, A., Chaban, A., Kit, I., & Len-diuk, T. (2022). Concept of the Intelligent Guide with AR Support. International Jour-nal of Computing, 21(2), 271-277. https://doi.org/10.47839/ijc.21.2.2596
[20] Lipianina-Honcharenko, K., Sachenko, A., Semaniuk, V., Badasian, A., & Kopania, Ł. (2023, September). Intelligent Method for Selecting Business Location in Smart City. In 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Ad-vanced Computing Systems: Technology and Applications (IDAACS) (Vol. 1, pp. 1184-1188).