UDC 004.58 + 004.93
DOI: 10.15587/2706-5448.2026.356218

**Olha Kryazhych,
Vasyl Vasenko,
Oleksandr Vasenko,
Anastasiia Pavlenko,
Kateryna Iushchenko**

DEVELOPMENT OF AN APPROACH TO FORMING A FRAME-BASED DICTIONARY FOR THE PERSONALIZATION OF AN EDUCATIONAL CHATBOT

The object of research is the process of functioning and using AI-based educational chatbots in the educational environment. The problem addressed in the research is the training of chatbots using dictionaries. The paper presents an approach to chatbot personalization through the use of a thematic dictionary and query adjustment with the help of prompts.

A frame-based model of a dictionary has been developed, which can be added to a chatbot as a PDF document. The frames represent the subject domain as a hierarchically organized system. User prompts and the processing of frame structures are integrated with the chatbot through thematic representations. This ensures flexible query formulation, scalability of dictionary resources, and the possibility of further expansion of the subject domain without violating the integrity of the language model.

Schemes for combining contextual projection of query interaction and dictionary search have been developed and substantiated. To implement prompts, an algorithm was designed based on the principle of a "marked bullet" selection of a term or expression. Chatbot personalization is achieved through the formation of a series of user-generated prompts.

Based on the results of experiments on adapting ChatGPT to users' educational needs, frame-based dictionaries were implemented and tested. For the sequential dictionary implementation scheme, at an accuracy level of 10^{-6}, the total computational complexity is approximately 4.67, while increasing the accuracy requirement to 10^{-8} reduces this value to 3.287. The hierarchical scheme, based on the frame organization of the dictionary and the use of TemaView, demonstrates comparable or lower complexity values (10^{-6} = 6.69 and 10^{-8} = 4.69).

The practical application lies in supporting the educational process through the use of personalized educational chatbots within learning systems.

Keywords: *prompt engineering, personalized learning, flexible query construction, table of thematic representations.*

Received: 10.01.2026
Received in revised form: 06.03.2026
Accepted: 29.03.2026
Published: 30.04.2026

How to cite
Kryazhych, O., Vasenko, V., Vasenko, O., Pavlenko, A., Iushchenko, K. (2026). Development of an approach to forming a frame-based dictionary for the personalization of an educational chatbot. Technology Audit and Production Reserves, 2 (2 (88)), 6–14. https://doi.org/10.15587/2706-5448.2026.356218

1. Introduction

A frame dictionary has a specific structure in which the meaning of a word or concept is described through a frame [1]. A frame is a typical situation (scenario) with a set of roles (slots) and their possible values [2]. Research [3] proves the need for additional training of large language models (LLM) on FrameNet annotations, which allows to increase the accuracy of solving problems similar to those presented in [4], especially in conditions of using new or specialized data (out-of-domain data).

The practical value of using frame dictionaries in additional training of LLM is to increase the accuracy of semantic interpretation of the text, reduce the ambiguity of lexical units and reduce the number of incorrect answers in specialized tasks. Due to the explicit representation of roles and relationships between entities, the model is able to more correctly identify the meaning of terms in context and generate more relevant answers. This is especially important for tasks similar to [4], where accurate interpretation of terms within a specific subject area is required.

Large language models themselves partially reflect frame semantics, forming consistent latent semantic structures that correspond to roles and event schemas [5]. However, the use of an explicit frame dictionary allows not only to limit the search for an answer to the subject area, but also to ensure controllability of interpretation, consistency of answers within the domain, and the possibility of explaining the obtained results through the frame structure. In addition, it contributes to the correct identification of the semantic frame for a word in context [6] and allows to operate with lexical meanings and cognitive loads represented in the frame structure.

Modern studies [7, 8] and reviews [9] confirm the need to adapt and personalize chatbots for educational and highly specialized tasks. In particular, in [9], "domain-specific fine-tuning" is clearly highlighted as a key practice for ensuring adequate and reliable answers. In [10], it is emphasized that chatbots with artificial intelligence (AI), trained on LLM, are promising digital tools for the educational process. The use of frame dictionaries in the process of additional training allows to improve the quality of educational content, reduce the risk of forming false knowledge and ensure greater compliance of answers with educational goals, which directly determines the effectiveness of their practical application. One of the conceptual foundations of the use of (AI) in education is the paradigm of a person in the decision-making loop [11]. Accordingly, intellectualized systems do not act as a replacement for a teacher or mentor, but as a tool to support learning. Within the framework of this approach, the ability to control the process of generating chatbot answers, as well as ensuring transparency and pedagogical validity of educational content, becomes of particular importance. A promising direction for solving these problems is the combination of hint engineering with the use of knowledge-based systems [12]. The current direction is the use of mechanisms for controlling the generation of answers based on hints in combination with external formalized sources of knowledge [13, 14].

In particular, in [15], a review of the use of chatbots with AI in higher education is carried out with an emphasis on personalized learning with content adaptation. At the same time, personalization is considered at the conceptual level without mechanisms for its implementation. Approaches based on specialized dictionaries and guided prompts are not analyzed. Therefore, practical implementation at the user level of a chatbot with generative AI is impossible.

In [16], an approach to creating an educational chatbot with adaptive functions is proposed. Personalization of learning occurs based on testing results and user behavior. The main emphasis is placed on algorithmic adaptation of learning scenarios. However, personalization through dictionaries or prompt templates is not considered, as the system uses predefined pedagogical rules.

In [17], the methodology for creating a personalized learning environment with an AI-based chatbot is considered. However, personalization tools (dictionaries, prompts, terminology management) are not considered.

In [18], personalization is interpreted as adapting content to learning progress. The mechanisms for personalizing language interaction through dictionaries and prompts are not analyzed due to different principles for building the analyzed chatbots.

In [19], the Selective Prompting Tuning method is proposed for personalized dialogues with LLM by selecting relevant prompts. At the same time, the educational context and the use of subject dictionaries are not considered. The research is of a general-model nature without specifying the technology for implementing the method.

In [20], an approach to personalized response generation using adaptive queries is presented. However, dictionaries as a separate pedagogical tool for the chatbot action scenario are not analyzed. The focus of the work is on general-purpose dialogue systems. Practical implementation at the user level is possible, but inefficient.

In [21], the creation of personalized prompts for a chatbot with generative AI is considered. However, the main emphasis of the work is on the creation of methodological recommendations for the development of educational content using chatbots with AI.

The analyzed works demonstrate a growing interest in the personalization of chatbots in education. At the same time, none of the studies offers a formalized approach to combining specialized dictionaries and structured prompts. As follows from the analysis of sources, there is a question of implementing a technology that will allow using chatbots in a personalized way at the level of an individual user. This determines the scientific need and justifies the relevance of further research.

The aim of research is to develop an approach to forming a frame dictionary for personalizing a chatbot based on generative artificial intelligence, which is accessed based on guided prompts. The LLM use in education is accompanied by a number of limitations associated with the variability of terminology and the lack of adaptation to the level of user training. For educational scenarios, such shortcomings are critical, as they can lead to the formation of incorrect ideas. This necessitates the transition from autonomous generative solutions to approaches focused on direct human-intelligent system interaction based on hint engineering.

To achieve the aim, the following objectives were set:

– to build a frame model of a dictionary that can be added to a chatbot with generative AI as a PDF document;

– to provide schemes for combining contextual search projection and query word comparison depending on the construction of user hints;

– to provide an algorithm for implementing the formation of hints for chatbot personalization;

– to experimentally confirm the effectiveness of the approach to chatbot personalization based on the use of a frame dictionary.

Thus, solving the problem of chatbot personalization using a frame dictionary and special hints is relevant. This will allow for the coordination of the user's work, taking into account their cognitive characteristics and the learning speed of the AI model to obtain a personalized assistant-mentor.

2. Materials and Methods

The object of research is the process of functioning and use of educational chatbots based on artificial intelligence in the educational environment.

The subject of research is personalization of the work of an educational chatbot by using a frame dictionary with response generation controlled by hint engineering.

The main hypothesis of research: the effectiveness of chatbot personalization for educational needs depends on the formalization of the thematic dictionary as a controlled semantic resource. Control and interaction with an AI-based chatbot occurs using prompts – a text query or instructions (prompts). In this case, there is no intervention in the parameters of the language model.

The work is simplified by the fact that a specialized dictionary is considered in the form of a PDF document. It contains structured dictionary articles with definitions, contexts of use or recommendations for the pedagogically correct use of terms. Such a dictionary is used as an external semantic resource when forming prompts for an educational chatbot, which allows for personalized learning. Such simplification ensures reproducibility of results, methodological transparency and the possibility of adapting the chatbot to different educational scenarios personally for a specific user. This is due to two components of research – a person in the decision-making loop and a knowledge-based system. For a human user, a PDF document simplifies the implementation of interaction with the chatbot. It also allows for formalization of hint engineering processes based on a combination of terms from dictionary tables.

The user implements the task of personalized learning using a chatbot, which it uses as a search or reference system for a specific educational topic. This is due to the fact that most existing chatbots are focused on supporting the fulfillment of individual educational needs, rather than on comprehensive personalization of learning (Table 1).

Based on Table 1, only some solutions (primarily ChatGPT and Claude) have sufficient functional potential to implement pedagogically meaningful personalization. Their key advantage is the ability to control the generation of answers using structured prompts and the use of specialized dictionaries, which allows to adapt the terminology, style of explanation and level of complexity. In addition, ChatGPT allows the user to define the didactic role of the chatbot (tutor, consultant, examiner). At the same time, other universal chatbots, focused mainly on search or office support, demonstrate limited suitability for educational scenarios. They do not support deep semantic personalization and control of the pedagogical context. Therefore, ChatGPT 5.2 Plus (Instant) was chosen for the research. The use of the frame method of presenting knowledge is due to the peculiarities of displaying stereotypical educational situations. The frame consists of characteristic cells. Therefore, the frame can be presented either in a complex way – in the form of ontologies, or simplified, as a table or structured linked text. For example, the frame "geometry" will contain cells: square, triangle, circle, trapezoid. There are sample frames (prototypes) and instance frames (specific situations).

 ISSN-L 2664-9969; E-ISSN 2706-5448

Table 1

Comparative characteristics of chatbots for educational needs

Universal chatbot	Model type	Educational opportunities	Personalization support	Working with dictionaries	Working with prompts	Pedagogical potential
ChatGPT	Generative LLM	Explanation of material, tutoring, text analysis, testing	High	Yes (via built-in or external dictionaries)	High (role, style, difficulty level)	Very high
Claude	Generative LLM	Explanation, working with large texts, academic writing	Medium-high	Limited	High (contextual prompts)	High
Gemini	Multimodal LLM	Explanation, interdisciplinary responses, STEM	Medium	Limited	Medium	Medium-high
Microsoft Copilot	LLM + office services	Support of educational materials, working with documents	Medium	No	Medium	Medium
Perplexity AI	LLM + search	Search and explanation of sources, literature reviews	Low-medium	No	Low	Limited
YouChat	LLM + search	Brief explanations, reference	Low	No	Low	Low

To solve the problems set, it is assumed that the dictionary is a register that has q^n cells of unit length. They are located on a straight line with coordinates $0 \div q^n - 1$ (q – the alphabet of the formation of dictionary words, n – the number of dictionary words $a_1 \ldots a_n$). The ordering is carried out by $q = 26$ – letters of the English alphabet, without taking into account the case. The number of dictionary words $n = 4000$ (terms without explanations). Part of the cells in the amount of N are active – for them the combination of words $a_1 \ldots a_n$ corresponds to the user's request for a specific educational need.

In this research, it is important to distinguish between a domain frame model and a dictionary frame model. A domain frame model describes a generalized knowledge structure independent of the user. A dictionary frame model is its applied interpretation, focused on a specific user or group of users. Accordingly, the same domain can be represented by a set of dictionary frame models, each of which reflects the individual learning trajectory, level of training and cognitive characteristics of the user.

In this experiment, the labor intensity H is understood as the average number of elementary actions required to complete the personalization process, in particular, entering hints to obtain the correct and most accurate answer.

The research was conducted using universal personal computers. The hardware consisted of an Intel Core i7 central processor (USA), 32 GB of DDR4 RAM, an NVIDIA GeForce RTX 3060 graphics adapter with 12 GB of video memory (USA), and a 1 TB NVMe SSD solid-state drive. This configuration provided the necessary computing resources to form a specialized dictionary and perform experimental studies using language models.

The Windows 11 Pro operating system (Microsoft, USA) was used as system software. The generalization and analysis of the experimental test results were carried out using MS Excel. The main programming language within the research was Python 3.10 (USA). NumPy 1.24, Pandas 2.0, and spaCy 3.7 libraries (USA, Germany) were used to implement algorithms and process data. An experimental assessment of the effectiveness of the specialized dictionary was performed using large language models via the OpenAI API (USA). The development and verification of the correctness of dictionary resources were carried out in the Visual Studio Code 1.85 environment (USA).

3. Results and Discussion

3.1. Frame model of the dictionary

The frame model of the dictionary (F-model) can be represented by text or a table. The distribution of individual cells (words) can be random or systematic. For the values N, q, n, the condition $N > 1$, $q^n > N$ must be satisfied. This implies that the probability r that an arbitrary cell of the dictionary with parameters N, q, n is used in the query is equal to N/q^n. In this case, $r < 1$, and the complexity H for forming the query and obtaining an adequate answer is minimal. In this case, the parameters N, q, n may not fall under the possible choice of the user. In this case, hints A_1, ... A_N are created, which help to find in the dictionary the corresponding combinations of words $a_1 \ldots a_n$, according to which a clarification is formed regarding the search for information on the query using the dictionary page m.

Personalization of an educational chatbot can be represented not only on the basic frame model of knowledge representation, but also on the construction of individualized frame models of dictionaries, which are adapted subsets of this model. Such dictionaries are formed taking into account the level of training of the user, its educational goals and the context of use, which allows to control the process of generating chatbot responses in a personalized mode.

The basic frame model of the subject area can be represented as a set

$$F\text{-}model = \{F_i, Slot_{ik}, Rel\}, \quad (1)$$

where F_i – frames corresponding to separate thematic sections of the subject area and may include one or more concepts; $Slot_{ik}$ – frame slots that reflect the terms of the corresponding sections and can acquire certain values; Rel – a set of relations between frames that define their semantic and hierarchical connections.

The frame model of the dictionary can be formed as a subset and projection of this model:

- by selected frames (topics);
- by a subset of slots (terms);
- taking into account the specified values and user restrictions.

Thus, the F-model of the dictionary is an adapted configuration of the basic model of the subject area, which is used directly in interaction with the chatbot.

Within the frame model, frames play the role of entities, and slots play the role of attributes, which allows to abandon the classical relational interpretation of objects and characteristics in favor of a more flexible and semantically rich knowledge structure. This is especially important for educational chatbots, where it is necessary to take into account the context, level of training and learning goals of the user.

Frames and slots in the model should be ordered according to their importance (priority) indicators, which are determined by experts on a discrete scale of 1, 2, 3, ..., n. The root element of the structure will be the main frame, which corresponds to the basic thematic section of the subject area. Other frames are placed hierarchically depending on the increase in their weight coefficients. A similar ordering principle is applied to slots within each frame, which will ensure an unambiguous place for each term in the overall structure.

As a result, a tree of the importance of the sections of the subject area and their terms will be formed. The root frame will reflect the key educational topic. Such a formation of the sections of the subject area can be presented as a table, in the form of a PDF document. PDF will record a single controlled version of knowledge. PDF document is a platform-independent standard. Such a structure becomes appropriate as a basis for personalized user interaction with an educational chatbot.

When working with the system interface, the user gains access to an importance tree from which it can select the necessary thematic sections and terms of the subject area, set the values of individual slots and form query conditions. Based on this data, the chatbot accesses the PDF dictionary document and generates a personalized result. For example, these can be the values of the searched terms or educational explanations adapted to the given context.

It is proposed to combine user prompts and frame structure processing in working with the chatbot using virtual tables – thematic representations (TemaView). This will ensure flexible query formation, scalability of dictionary resources and the possibility of further expansion of the subject area without violating the LLM integrity. The scheme of interaction of frame dictionary objects and user prompts is shown in Fig. 1.

The ThemeView will be processed not as a regular table, but as a contextual projection of the frame model of the dictionary (and not the entire subject area). Each ThemeView will correspond to a specific dictionary configuration formed for the current user or educational task. The general processing logic is as follows:

1) selecting a topic/subtopic from the PDF dictionary document, while activating the corresponding ThemeView;

2) filtering frames and slots by priorities, user level, learning goals;

3) filling slots with values from the PDF dictionary document or from the current query;

4) context formation – a structured system of prompts for generating a response by the chatbot.

Educational priorities may require the use of a number of dictionaries of thematic representations of individual areas. Each thematic representation corresponds to a separate frame of the subject domain frame model, and its attributes are interpreted as slots of the corresponding frame. The content of TemaView is formed dynamically based on information stored in dictionaries. Such a mechanism will ensure the relevance and integrity of the knowledge used by the educational chatbot in the process of generating personalized answers. In addition, such a dictionary can be formed individually by a teacher (lecturer, tutor) with subsequent transfer to students for personalization of the educational chatbot. Thus, the subject domain frame model will act as the basic ontological structure of knowledge, while the dictionary frame model will be its applied implementation, which ensures personalization of interaction with the chatbot. The presence of a set of such dictionaries allows to adapt the system to different users without changing the basic model, which is a key condition for scalability and effective LLM use in the educational process.

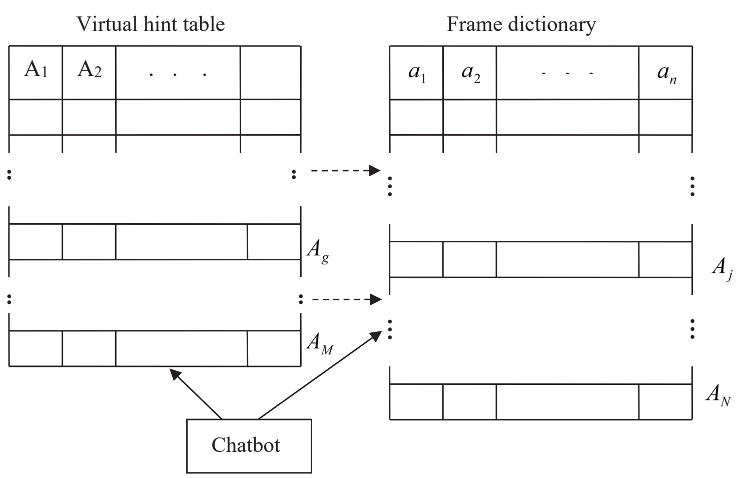

Fig. 1. Thematic representation of the interaction of frame dictionary objects and user prompts when personalizing a chatbot

3.2. Schemes for combining contextual projection of search and comparison of query words depending on the construction of hints

The creation of TemaView is carried out in accordance with predefined dictionary rules that regulate their identification, ordering and integration into the general structure of frames. Such rules include the naming rule, the numbering rule and the connection rule.

The naming rule involves assigning each TemaView a unique name, which allows it to be uniquely identified at the level of database objects, in particular tables, representations and procedures, on the basis of which the software implementation of the educational chatbot functions. In the name structure, it is proposed to use a fixed prefix, for example "TV", which consists of the first two or three words selected from the dictionary. The number of initial words can vary, because this does not affect the unification and convenience of processing.

The numbering rule consists in including a numerical identifier in the TemaView name, which is placed after the prefix. The value of this identifier corresponds to the importance indicator of the corresponding frame in the frame model. This approach allows to form a tree of the importance of thematic sections of the subject area and use it in the interface of the educational chatbot to control the depth and sequence of presentation of educational material.

After the prefix and numerical identifier, an arbitrary name should be used in the TemaView name, which reflects the content of the corresponding thematic section. A short name can be used for convenience of work at the level of application logic, while the full name of the representation is proposed to be used directly in the dictionary. Such a naming structure contributes to the transparency of the architecture of the frame dictionary and simplifies the integration of thematic representations into the personalization mechanism of the educational chatbot.

There are two possible schemes for combining TemaView when interacting with objects for personalizing the chatbot – sequential and hierarchical. The sequential TemaView (TV) scheme is a network by the ID key (Fig. 2), for example, according to the description of page *m* of the dictionary.

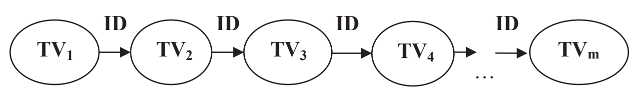

Fig. 2. Sequential scheme of combining TemaView

As a result, when combining all TemaViews, the relation Q is obtained, which can be represented in disjunctive form

$$Q = TV_1 \cdot TV_2 \cup TV_2 \cdot TV_3 \cup TV_3 \cdot TV_4 \cup \ldots \cup TV_{m-1} \cdot TV_m = \\ = U_2^m TV_{i-1} \cdot TV_i, \quad (2)$$

where $i = 2, \ldots, m$. This is due to the fact that the operation is performed on pairs of adjacent thematic representations; $TV_{i-1} \cdot TV_i$ – a conjunction that holds tuples formed as a result of the Cartesian product of relations that satisfy the equality condition for the common key ID; U – the formation of a combination of words on a dictionary page.

The disjunction is formed as a result of the application of the left outer join operation, within which the resulting relation includes tuples with unspecified attributes of the right parts of the conjunctions (with NULL values). During the implementation of model (2), the user's query deepens from simple to complex. Clarifications are formed gradually, and the question receives a detailed answer taking into account each subsequent hint.

The disadvantage of this scheme is that the absence of data in any of the TemaViews leads to a break in the chain and the inaccessibility of its further right part. That is, data from all TemaViews located to the right will not be selected.

To select any or all tuples using dictionary data, it is worth forming hints to create a hierarchical scheme of combining TemaViews. In such a scheme, all TemaViews are combined with one main TemaView by the key ID. This scheme will be a directed graph G of the "fan" type, consisting of a main root vertex of degree $m-1$ and $m-1$ hanging vertices (leaves) of the first degree (Fig. 3).

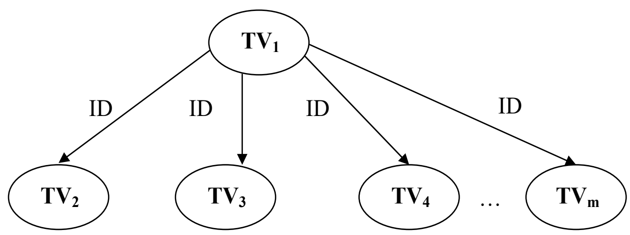

Fig. 3. Hierarchical scheme of TemaView connection

Graph G is a bipartite graph that models the relationship between the main TemaView – TV_1 (root frame, common section of the dictionary) and other TemaView (frames, thematic sections). If the dictionary contains data, then TV_1 will also be non-empty. The vertex cover (the minimum number of vertices connected to all other vertices) of such a graph is 1. Therefore, it can be argued that it is the optimal option for representing the structure of a frame dictionary, since it provides the possibility of relatively simple and shortest access to the data of all TemaViews connected to the common vertex TV_1. The edges of graph G are the common key ID, which is used to connect all TemaViews.

Adding a new frame or removing an existing one in the user interface will be interpreted as adding a new vertex to graph G or removing an existing one. The result of such a connection is a relation in disjunctive normal form

$$Q = TV_1 \cdot TV_2 \cup TV_1 \cdot TV_3 \cup TV_1 \cdot TV_4 \cup \ldots \cup TV_1 \cdot TV_m = \\ = U_2^m TV_1 \cdot TV_i, \qquad (3)$$

where $i = 2, \ldots, m$.

The disjunction will be formed as a result of the application of the left outer join operation, within which the resulting relation includes Q tuples with unspecified attributes of the right parts of the conjunctions $TV_1 \cdot TV_i$ (NULL values). The table of such a dictionary will contain the main term and subordinate or related terms by a common attribute (regarding the subject, topic, object of research).

3.3. Algorithm for generating hints for chatbot personalization

The proposed framework model of knowledge representation and the thematic representation mechanism create a formalized basis for implementing hint engineering for personalizing an educational chatbot. Within this approach, TemaView plays the role of an intermediate semantic layer between the structured knowledge of the subject area and the generative language model, providing controlled context formation for generating answers.

Each TemaView, corresponding to a separate frame of the frame model, aggregates relevant slots and their values, formed taking into account the priority of frames, user parameters and the current educational query. The result of TemaView processing will be a structured topic context, which is directly used to build a prompt. Unlike static or universal prompts, such a prompt will be dynamic in nature and is formed automatically based on the current state of the frame dictionary. The graph organization of TemaView in the form of a bipartite graph with the root vertex TV_1 (Fig. 3) will provide effective management of the complexity of the prompt. Thanks to the use of a common ID key and left outer join operations, it becomes possible to form disjunctive contexts in which the absence of data in individual frames does not block access to other thematic representations. This will allow to gradually expand or refine the prompt during the dialogue, implementing the principle of gradual deepening of the query and even expanding or deepening the topic taking into account the TV_1 probability when forming the user's query.

The process of forming a query based on a frame dictionary can be defined as displaying data at the input, processing it, and presenting data at the output.

The input will include:
– initial thematic representation (base word, expression, or set of terms);
– a set of possible dictionary cells;
– search parameters (permissible error probability, user level);
– current query context.

The processing process will include:
– iterative selection of options ("ballbox");
– application of user prompts;
– filtering options according to the frame structure;
– checking the conditions for completing the process.

Then the output may be an identified term or set of terms, a message about the absence of a result, or a refined user query.

Provided that TV_1 is available in the dictionary (a base word, expression, a selection of terms), this process is expedient to describe as a formalized search process with defined input data, processing rules and execution results. This can be implemented in the form of a "tagged ball" algorithm:

Step 1. Check the TV_1 presence in the dictionary. If TV_1 is available, then the deterministic ("tagged") ball is previously placed in the urn and the transition to Step 2 is made.

If TV_1 is absent, then the "tagged" ball is not added, and the process also proceeds to Step 2.

Step 2. At iteration and, a random selection of balls from the set of possible options (cells of the dictionary) is performed. The number of selected balls q is determined.

Step 3. If $q = 0$, i. e. no ball was selected (the corresponding cell of the dictionary is absent), then:
– in the case when the "tagged" ball was previously added, the process is terminated;
– otherwise, return to Step 1 to refine the query.

Step 4. If $q > 1$, i. e. more than one ball is selected, the process continues: the user is given a hint to refine the query, after which the transition to Step 2 is made.

Step 5. If $q = 1$, i. e. exactly one ball is selected, then:
– if it corresponds to a "marked" ball, the process ends with a correct result;
– if it is "falsely marked", the process ends with a false positive result, and a new user query is formed.

Step 6. Completion of the algorithm.

The proposed algorithm will provide a procedural implementation of the specified process and allow its formalized execution within the user's interaction with the chatbot. When personalizing a chatbot using a dictionary with hints, there are three possible scenarios. In this case, the complexity of the personalization work H and its calculated components are taken into account with a superscript "3". This can be written as follows:

$$H_{13}^{(3)} = H_{23}^{(3)} = \left(\mu_2 + k\right);$$
$$\bar{v}_1^{(3)} = \sum_{i=1}^{n} i \cdot p_{1i}^{(3)} \prod_{s=1}^{i-1}\left(1 - p_{1i_s}^{(3)}\right);\ \bar{v}_2^{(3)} = \sum_{i=1}^{n} i \cdot P_{2i}^{(3)} \prod_{s=1}^{i-1}\left(1 - P_{2i_s}^{(3)}\right);$$
$$p_{1i}^{(3)} = P\left(o, r, q^{n-i}\right);\ p_{2i}^{(3)} = \sum_{g=0}^{1} P\left(g, r, q^{n-i}\right);$$
$$\bar{m}_1^{(3)} = 1;\ \bar{m}_2^{(3)} = \sum_{i=1}^{n} 1 \cdot P\left(1, r, q^{n-i}\right) \cdot \prod_{s=1}^{i-1}\left(1 - p_{2i_s}^{(3)}\right). \qquad (4)$$

In model (4) for $\bar{m}_2$ the third group factor will determine the probability that the process did not end at steps 1, 2, ... $i-1$. The second factor will determine the conditional probability that the process ended at step and false positive, and the first – the number of words analyzed in this case.

3.4. Experimental results and their discussion

The TemaView processing process can be initiated by selecting a topic or subtopic of the subject area of the dictionary in accordance with the user's educational request. At this stage, the corresponding thematic representation is activated, within which the selection of frames is carried out taking into account their priority, determined by expert means when developing such a dictionary. Frame priorities allow to adjust the depth and complexity of the educational material, which is an important factor in personalizing educational content for users with different levels of training.

Further processing of TemaView involves filtering slots inside the selected frames. Slots are considered as carriers of terms and parameters of the subject area, the values of which can be explicitly specified by the user or automatically determined based on its profile, interaction history and current query context. This ensures adaptation of educational material to individual educational needs, learning style and goals.

As a result of processing the thematic representation, a structured topic context is formed, containing an ordered set of frames, slots and their current values. This context is used as an intermediate model between the frame dictionary and the response generation module of the educational chatbot. Thanks to this approach, the amount of irrelevant information is reduced, the semantic consistency of responses is increased and the risk of incorrect generation of educational content is reduced.

From a technical point of view, TemaView processing can be implemented both at the level of the database management system in the form of virtual representations and at the level of the application logic of the chatbot. Within the framework of the applied implementation of TemaView, it is advisable to present it in the form of data structures suitable for further processing and integration with language models, which ensures scalability and extensibility of the frame dictionary without violating the integrity of the general knowledge model.

The structured topic context formed on the basis of TemaView is used to build a controlled input query to the language model, which allows directing the process of generating answers within a given subject area. This approach provides a combination of formalized knowledge of the frame dictionary with the capabilities of generative AI and creates the basis for the implementation of a personalized educational chatbot with predicted behavior and pedagogically justified results.

Tables 2 and 3 show the results of the experiment on the personalization of the chatbot based on the formed frame dictionary.

Table 2 presents the results of experimental evaluation of the use of a sequential dictionary formation scheme in chatbot personalization. The indicators q of the number of variants (dictionary cells) corresponding to the current query at the kth step, the total number of cells n, the number of target cells N of the dictionary and the probability r of selecting a relevant cell at one step are considered. Columns with numbers 1–8 correspond to the sequential steps of forming clarifying hints. The corresponding cells contain the threshold values of the probability of error after each step, which are used to indirectly estimate the probability of completing the search process. The rows correspond to the specified threshold values of the probability of error (10^{-6} and 10^{-8}), which determine the permissible level of risk of misidentification of the term.

Thus, Table 2 shows how the distribution of the marginal error probabilities changes during the query refinement process, which indirectly characterizes the probability of completing the search in the chatbot adaptation process depending on the number of terms entered and the specified dictionary correspondence. This allows not only to algorithmically adapt training scenarios, similar to [16], but also to implement prompt templates specific to a particular user.

The analysis of the values shows that for small step numbers (1–4), the probability of completing the process is close to zero. This means that after entering only a few refinements, the dictionary still contains a significant number of variations and the answer may be inaccurate. Starting from steps 5–7, there is an increase in the values of the marginal error probability, which corresponds to the transition to the zone of choice certainty and the actual completion of the search process. This indicates a significant narrowing of the set of possible terms, which deepens the results presented in the works [17, 19].

In particular, at the 6th–7th step, there is a concentration of values that meet the boundary conditions for the completion of the process (values of the order of 0.35–0.71 and 0.28–0.63, depending on the parameters). This means that it is at this stage that the typical completion of the search in the sequential scheme occurs and the chatbot is personalized to the user's needs. The probabilities at the 8th step and further decrease significantly, which indicates a small number of cases when more extended prompts are required to uniquely identify the term.

Table 3 presents the results of experimental evaluation of the hierarchical chatbot personalization scheme based on the use of a frame dictionary and thematic representations. Unlike the sequential scheme, in this case the search and identification process of the term is carried out taking into account the hierarchical structure of the frames and their priorities, which allows to significantly reduce the search space.

Table 2

Probability distribution $\rho_1^3(i)$: sequential scheme

Dictionary	r	N	n							
			1	2	3	4	5	6	7	8
$q = 26$ $n = 4000$	10^{-6}	$1.1 \cdot 10^6$	0	0	$36e^{-15}$	0.35043	0.62862	0.021	$2e^{-5}$	$7e^{-10}$
	10^{-8}	$1.1 \cdot 10^4$	0	$2e^{-5}$	0.71493	0.28207	0.00297	$9.7e^{-7}$	$1e^{-11}$	$3e^{-18}$

Table 3

Probability distribution $\rho_1^3(i)$: hierarchical scheme

<table>
<tr><th rowspan="2">Dictionary</th><th rowspan="2">r</th><th rowspan="2">N</th><th colspan="8">n</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr>
<tr><td rowspan="6">$q = 26$
$n = 4000$</td><td>10^{-6}</td><td>$1.1 \cdot 10^6$</td><td>0</td><td>0</td><td>$5e^{-435}$</td><td>$4e^{-44}$</td><td>$4.5e^{-5}$</td><td>0.368</td><td>0.572</td><td>0.06</td></tr>
<tr><td>10^{-8}</td><td>$1.1 \cdot 10^4$</td><td>$5e^{-435}$</td><td>$4e^{-44}$</td><td>$4.5e^{-5}$</td><td>0.3678</td><td>0.572</td><td>0.059</td><td>0.001</td><td>$6e^{-7}$</td></tr>
<tr><td rowspan="2">r</td><td rowspan="2">N</td><td colspan="8">n</td></tr>
<tr><td colspan="2">9</td><td colspan="2">10</td><td colspan="2">11</td><td colspan="2">12</td></tr>
<tr><td>10^{-6}</td><td>$1.1 \cdot 10^6$</td><td colspan="2">0.0006</td><td colspan="2">$6e^{-7}$</td><td colspan="2">$5.9e^{-11}$</td><td colspan="2">$6e^{-16}$</td></tr>
<tr><td>10^{-8}</td><td>$1.1 \cdot 10^4$</td><td colspan="2">$6e^{-11}$</td><td colspan="2">$6e^{-16}$</td><td colspan="2">$5.98e^{-22}$</td><td colspan="2">$5.98e^{-29}$</td></tr>
</table>

As in the previous experiment (Table 2), column numbers 1–12 correspond to the steps of the prompt input process, after which the threshold values of the error probability are determined. The table presents the results for two levels of the permissible error probability – 10^{-6} and 10^{-8}.

The analysis of numerical values shows that non-zero values of the threshold error probability appear already at the 4th–5th steps of the hierarchical scheme. In the sequential scheme (Table 2), the probabilities at these steps were practically equal to zero. This indicates that the use of the frame hierarchy allows to narrow down the set of possible terms earlier due to semantic selection. This allows the implementation of a similar approach at the user level, in contrast to the one proposed in [20].

The maximum values of the indicators corresponding to the threshold conditions for the completion of the process in the hierarchical scheme are observed at the 6th–7th steps. Here, the probability values reach approximately 0.37–0.57, depending on the given level of accuracy. At the same time, a significant proportion of the probability is also preserved at the 5th step, which was practically not observed in the sequential scheme. This means that in a significant number of cases, chatbot personalization can be achieved after entering a smaller number of prompts. Extending the process to the 9th–12th steps is accompanied by a decrease in the values of the marginal probability of error (of the order of 10^{-11}–10^{-29}). This indicates a small number of extreme cases when the term has similarities with other dictionary elements. It should be noted that Table 2 and Table 3 directly show the marginal probabilities of error. The probability of completing the search process is interpreted on the basis of the indicated tables as a complementary value. Thus, the hierarchical scheme not only reduces the average length of the prompt, but also limits the number of situations with excessive labor intensity, in contrast to the results of work [16].

Table 4 presents the results of a quantitative assessment of the labor intensity of the chatbot personalization process for two options for the formation of a specialized dictionary – sequential and hierarchical schemes. The calculations were performed for two levels of permissible error probability (10^{-6} and 10^{-8}), which meet different requirements for the accuracy of term identification.

As a result, the overall user interaction with the chatbot becomes more manageable and predictable, which is fundamentally important for the implementation of personalization based on prompts and the dictionary.

Thus, Table 4 confirms that the use of a hierarchical frame dictionary leads to a redistribution of its components in favor of semantically significant stages, which proves the effectiveness of the approach in contrast to [15]. This provides more effective personalization of the educational chatbot, reduces the cognitive load on the user and increases the quality of the formed responses.

The practical significance of research lies in improving the interaction between the educational chatbot and the user by implementing personalization mechanisms based on the use of a frame dictionary and thematic representations. The proposed approach ensures the adaptation of educational content to individual educational needs, the level of training and the context of the user's request, which contributes to increasing the efficiency of learning the educational material. The practical value of the research results also lies in reducing the complexity of the user's interaction with the chatbot, reducing the number of clarifying requests and increasing the predictability of the system's behavior.

A limitation of research is the dependence of the effectiveness of personalization of the educational chatbot on the completeness and quality of the formed frame dictionary, as well as on the expert definition of the structure of frames and their priorities. An additional limitation is the use of an experimental model of interaction with language models through hint engineering mechanisms in query management, which does not allow to fully assess the influence of the internal parameters of the language model on the results of personalization.

Under martial law, the specified research can be used to ensure the continuity of the educational process by implementing personalized educational chatbots in distance and blended learning systems. The proposed approach allows adapting educational content to the individual needs of students in conditions of limited access to traditional forms of education, unstable schedule and increased psychological stress. The practical application of the research results also lies in the possibility of operational support for self-study, in particular in computer science disciplines, through personalized explanation of the material and step-by-step user support.

Prospects for further research lie in integrating the proposed approach with a model of human behavior in the decision-making loop. This will allow combining automated personalization with pedagogical control and expert validation of educational content. An important task is to analyze the impact of the structure of the frame dictionary on the quality and stability of response generation by language models.

Table 4

The value of labor intensity $H^{(3)}$ and its components

Vocabulary formation scheme	r	N	$H^{(3)}$		$\overline{v}_1^{(3)}$	$\overline{m}_1^{(3)}$	$\overline{v}_2^{(3)}$	$\overline{m}_2^{(3)}$
			$\delta = 1.0$	$\delta = 0.75$				
Sequential	10^{-6}	$1.1 \cdot 10^6$	2.885	2.795	4.67	1	4.67	0
	10^{-8}	$1.1 \cdot 10^4$	2.256	2.166	3.287	1	3.287	0
Hierarchical	10^{-6}	$1 \cdot 10^6$	6.031	5.9	6.69	1	6.69	0
	10^{-8}	$1 \cdot 10^4$	4.543	4.413	4.69	1	4.69	0

For a sequential scheme with an accuracy level of 10^{-6}, the total value of the complexity is $H \approx 4.67$, while for a more stringent criterion of 10^{-8} it decreases to $H \approx 3.287$. This decrease is explained by the fact that with increasing accuracy requirements, the probability of early negative or unambiguous completion of the process increases, which reduces the average number of analyzed steps.

For a hierarchical scheme that uses the dictionary frame structure and TemaView, the complexity values are consistently lower or comparable, despite the greater complexity of the initial selection. In particular, at an accuracy level of 10^{-6}, the value of H is 6.69, and at 10^{-8} – 4.69. At the same time, the share of actions related to semantic narrowing of the search space increases, which reduces the number of iterations in the dialog with the user.

Individual components of the complexity, given in Table 4, demonstrate that in the hierarchical scheme most of the costs fall on the initial stage of context formation (request and prompt), but this is compensated by a reduction in the number of subsequent clarifications.

4. Conclusions

1. A frame model of the dictionary is presented as an adapted projection of the frame model of the subject area, focused on a specific user or educational task. Such a dictionary can be integrated into a chatbot with generative AI in the form of a PDF document. Unlike traditional dictionaries, the proposed approach allows to represent knowledge in the form of a system of interconnected frames, slots and relations between them, which provides a semantically rich and structured presentation of information. This creates the possibility of forming a set of dictionary variants for one subject area depending on educational needs, level of training and context of use, which increases the effectiveness of chatbot personalization.

2. Schemes of combining contextual projection of search and comparison of query words depending on the construction of user prompts are developed. Two schemes are presented and justified – sequential and

hierarchical. The advantage of a hierarchical scheme for building a frame dictionary is proven. This allows to build a dictionary by thematic sections. As a result, such a scheme allows to have a structured dictionary divided into topics, which greatly simplifies the personalization process.

3. An algorithm for implementing the formation of prompts for chatbot personalization is presented. The algorithm is based on the approach of probability theory when solving the "marked bullet" problem. This allows to search for information with prompts and form it in response according to personalized features. As a result, a structured topic context is formed with an ordered set of frames, slots and their current values.

4. The experiments conducted confirmed the effectiveness of the approach to chatbot personalization based on the use of a frame dictionary. When forming a frame dictionary according to a sequential scheme, the effectiveness of personalization is observed closer to the 8th prompt. When building a frame dictionary according to a hierarchical scheme, non-zero probabilities of completing the personalization process appear at Steps 4–5. For a sequential dictionary construction scheme with an accuracy level of 10^{-6}, the total complexity is about 4.67, while with increasing accuracy requirements to 10^{-8} it decreases to 3.287. The hierarchical scheme, based on the frame structure of the dictionary and the use of thematic representations, demonstrates comparable or lower complexity indicators: at an accuracy of 10^{-6} – 6.69, and at 10^{-8} – 4.69.

Acknowledgments

The authors express their gratitude to the employees of the Educational and Methodological Laboratory of Digital Education and Artificial Intelligence of the Hryhorii Skovoroda University in Pereiaslav for their assistance in conducting a computational experiment and testing the developed approach.

Conflict of interest

The authors declare that they have no conflict of interest regarding this research, including financial, personal, authorship or other, that could influence the research and its results presented in this article.

Financing

The research was conducted without financial support.

Data availability

The manuscript has no related data.

Use of artificial intelligence

The authors used artificial intelligence technologies within the permissible framework to provide their own verified data, which is described in the research methodology section.

Authors' contributions

Olha Kryazhych: Conceptualization, Methodology, Writing – original draft, Writing – review and editing; ***Vasyl Vasenko***: Supervision, Data curation, Writing – review and editing; ***Oleksandr Vasenko***: Project administration, Resources, Validation, Software; ***Anastasiia Pavlenko***: Data curation, Formal analysis; ***Kateryna Iushchenko***: Investigation, Formal analysis.

References

1. Fillmore, C. J. (1982). Frame semantics. *Linguistics in the Morning Calm*. Hanshin Publishing Company, 111–137. Available at: https://brenocon.com/Fillmore%201982_2up.pdf
2. Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Scheffczyk, J. (2010). *FrameNet II: Extended theory and practice*. International Computer Science Institute. Available at: https://my.eng.utah.edu/~cs6961/papers/FrameNet_book.pdf
3. Chundru, J. K., Poddar, R., Cao, J., Jiang, T. (2025). Do LLMs Encode Frame Semantics? Evidence from Frame Identification. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Suzhou: Association for Computational Linguistics, 29476–29488. https://doi.org/10.18653/v1/2025.emnlp-main.1499
4. Kryazhych, O., Ivanov, I., Iushchenko, K., Kupri, O., Vasenko, O., Riznyk, V. et al. (2025). Devising an approach to preventing information chaos in chat bots using generative artificial intelligence. *Eastern-European Journal of Enterprise Technologies, 2 (2 (134))*, 84–95. https://doi.org/10.15587/1729-4061.2025.324957
5. Hase, P., Diab, M. T., Celikyilmaz, A., Li, X., Kozareva, Z., Stoyanov, V. et al. (2021). Do Language Models Have Beliefs? Methods for Detecting, Updating, and Visualizing Model Beliefs. *ArXiv*. Available at: https://www.semanticscholar.org/paper/Do-Language-Models-Have-Beliefs-Methods-for-and-Hase-Diab/4a247cbfca9dcf91e2da24e6d2d84601a9041a8f
6. Millière, R., Buckner, C. (2024). A philosophical introduction to language models – Part I: Continuity with classic debates. *arXiv*. https://doi.org/10.48550/arXiv.2401.03910
7. Li, Y., Qu, S., Shen, J., Min, S., Yu, Z. (2023). Curriculum-Driven Edubot: A Framework for Developing Language Learning Chatbots Through Synthesizing Conversational Data. *arXiv*. Available at: https://arxiv.org/abs/2309.16804
8. Bayraktar, R., Sarıtürk, B., Elmas Erdem, M. (2024). Assessing Fine-Tuning Efficacy in LLMs: A Case Study with Learning Guidance Chatbots. *International Journal of Innovative Science and Research Technology,* 2461–2471. https://doi.org/10.38124/ijisrt/ijisrt24may1600
9. Aldhafeeri, L., Aljumah, F., Thabyan, F., Alabbad, M., AlShahrani, S., Alanazi, F. et al. (2025). Generative AI Chatbots Across Domains: A Systematic Review. *Applied Sciences, 15 (20),* 11220. https://doi.org/10.3390/app152011220
10. McGrath, C., Farazouli, A., Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: a review of an emerging research area. *Higher Education, 89 (6),* 1533–1549. https://doi.org/10.1007/s10734-024-01288-w
11. Memarian, B., Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans, 2 (1),* 100053. https://doi.org/10.1016/j.chbah.2024.100053
12. Wölfel, M., Shirzad, M. B., Reich, A., Anderer, K. (2023). Knowledge-Based and Generative-AI-Driven Pedagogical Conversational Agents: A Comparative Study of Grice's Cooperative Principles and Trust. *Big Data and Cognitive Computing, 8 (1),* 2. https://doi.org/10.3390/bdcc8010002
13. Peng, B., Galley, M., He, P., Cheng, H., Xie, Y., Hu, Y. et al. (2023). Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback. *arXiv*. https://doi.org/10.48550/arXiv.2302.12813
14. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y. et al. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv*. https://doi.org/10.48550/arXiv.2312.10997
15. El mourabit, I., Jai Andaloussi, S., Ouchetto, O., Miyara, M. (2025). AI Chatbots in Higher Education: Opportunities and Challenges for Personalized and Mobile Learning. *International Journal of Interactive Mobile Technologies (IJIM), 19 (12),* 19–37. https://doi.org/10.3991/ijim.v19i12.54163
16. Chen, D.-L., Aaltonen, K., Lampela, H., Kujala, J. (2024). The Design and Implementation of an Educational Chatbot with Personalized Adaptive Learning Features for Project Management Training. *Technology, Knowledge and Learning, 30 (2),* 1047–1072. https://doi.org/10.1007/s10758-024-09807-5
17. Davar, N. F., Dewan, M. A. A., Zhang, X. (2025). AI Chatbots in Education: Challenges and Opportunities. *Information, 16 (3),* 235. https://doi.org/10.3390/info16030235
18. Labadze, L., Grigolia, M., Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education, 20 (1).* https://doi.org/10.1186/s41239-023-00426-1
19. Huang, Q., Liu, X., Ko, T., Wu, B., Wang, W., Zhang, Y. et al. (2024). Selective Prompting Tuning for Personalized Conversations with LLMs. *Findings of the Association for Computational Linguistics ACL 2024*, 16212–16226. https://doi.org/10.18653/v1/2024.findings-acl.959
20. Raj, K., Roy, K., Bonagiri, V., Govil, P., Thirunarayan, K., Goswami, R. et al. (2024). K-PERM: Personalized Response Generation Using Dynamic Knowledge Retrieval and Persona-Adaptive Queries. *Proceedings of the AAAI Symposium Series, 3 (1),* 219–226. https://doi.org/10.1609/aaaiss.v3i1.31203

ISSN-L 2664-9969; E-ISSN 2706-5448

21. Krouska, A., Troussas, C., Voyiatzis, I., Mylonas, P., Sgouropoulou, C. (2024). ChatGPT-based Recommendations for Personalized Content Creation and Instructional Design with a Tailored Prompt Generator. *2024 2nd International Conference on Foundation and Large Language Models (FLLM).* Dubai: IEEE, 295–299. https://doi.org/10.1109/fllm63129.2024.10852487

✉***Olha Kryazhych,*** *PhD, Senior Researcher, Associate Professor, Institute of Telecommunications and Global Information Space National Academy of Sciences of Ukraine, Kyiv, Ukraine, e-mail: economconsult@gmail.com, ORCID: https://orcid.org/0000-0003-1845-5014*

Vasyl Vasenko, *PhD, Associate Professor, Head of Department of Theory and Techniques of Technology Education and Computer Graphics, Hryhorii Skovoroda University in Pereiaslav, Pereiaslav, Ukraine, ORCID: https://orcid.org/0000-0002-2527-6359*

Oleksandr Vasenko, *PhD, Associate Professor, Head of Department of Digital Methods of Teaching, Hryhorii Skovoroda University in Pereiaslav, Pereiaslav, Ukraine, ORCID: https://orcid.org/0000-0002-8895-4284*

Anastasiia Pavlenko, *PhD, Associate Professor, Department of Digital Learning Technologies, Hryhorii Skovoroda University in Pereiaslav, Pereiaslav, Ukraine, ORCID: https://orcid.org/0000-0002-5863-0385*

Kateryna Iushchenko, *PhD, Junior Researcher, Institute of Telecommunications and Global Information Space of National Academy of Sciences of Ukraine, Kyiv, Ukraine, ORCID: https://orcid.org/0000-0001-5183-816X*

✉*Corresponding author*