

Development of Intelligent Chat-bots for Support of Scientific Activities and Postgraduate Education: Role of Large Language Models and Semantic Technologies★

Igor Sinitsyn^{1,†}, Pylyp Andon^{1,‡}, Julia Rogushina^{1,2,*,†}, Boris Lashchonov^{1,†}, Oleksandr Simonenko^{1,†}

¹ Institute of Software Systems, National Academy of Sciences of Ukraine, 44 Glushkov Pr., Kyiv, 03680, Ukraine,

² Institute for Digitalisation of Education, National Academy of Educational Sciences of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

Abstract

This paper presents the development of an intelligent assistant for postgraduate education that integrates large language models with semantic technologies, employing a hybrid architecture that combines flow-based and playbook-based agents. The system leverages Vertex AI Agent Builder and incorporates the Gemini model to generate natural language responses, while Semantic MediaWiki and domain ontologies ensure semantic consistency and explainability. The assistant supports multilingual interaction, personalized recommendations, and transparent reasoning through human-in-the-loop mechanisms. Challenges such as hallucinations, confidentiality risks, and integration with outdated institutional systems are addressed by introducing semantic markup, knowledge graphs, and retrieval-augmented generation. The paper also discusses the Linza platform as a technological foundation for complex experience extraction from natural language documents, highlighting its ability to transform heterogeneous unstructured texts into complex information objects, and analyze the place of intelligent chat-bots into the ecosystem of Linza.

Keywords

Intelligent Chat-bots, Large Language Models, Semantic Technologies, experience extraction, Human-in-the-Loop

1. Introduction

Now intelligent assistants based on chat-bots are important elements of modern information infrastructure. Their emergence is caused by the need to ensure rapid access to relevant knowledge, facilitate navigation across diverse documents, support the integration of information from multiple sources, and generate personalized responses for users while providing explanations of how these responses are constructed.

A chat-bot is a software system designed to simulate natural language (NL) communication. These systems usually employ generative artificial intelligence (AI) technologies and natural language processing (NLP) to understand and generate responses resembling human interaction. AI-based chat-bots utilize Large Language Models (LLMs) to produce more comprehensive and personalized answers. Intelligent assistants extend the functional capabilities of chat bots that provide fixed responses by means of enhanced language understanding, planning, and integration of external knowledge and tools.

2nd Workshop "Intelligent information technologies" UkrProg-IIT'2026 co-located with 16th International Scientific and Practical Programming Conference UkrPROG'2026, May 5-6, 2026, Kyiv, Ukraine

* Corresponding author.

† These authors contributed equally.

✉ ips2014@ukr.net (I. Sinitsyn); andon@isofts.kiev.ua (P. Andon); ladamandraka2010@gmail.com (J. Rogushina); snowbor1@gmail.com (B. Lashchonov); morden2k@gmail.com (O. Simonenko)

ORCID 0000-0002-4120-0784 (I. Sinitsyn); 0009-0000-1737-5579 (P. Andon); 0000-0001-7958-2557 (J. Rogushina); 0009-0006-2055-2058 (B. Lashchonov); 0009-0001-7443-9492 (O. Simonenko)



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

Such software systems are currently applied in various domains such as science, education [1], medicine, advisory services. Intelligent chat-bots are increasingly integrated into various stages of scientific activities, performing functions of administrative and informational support, as well as organizing research processes and training scientific personnel. Another important direction is the application of chat-bots in postgraduate education, directly linked to the learning of future researchers. Intelligent chat-bots assist postgraduate and doctoral students in navigating regulatory documents, curricula, and research methodologies, while also generating personalized sets of learning materials tailored to their research activities [2].

Chat-bots supporting postgraduate education enable:

- Navigation through specialty-specific and general regulatory documents;
- Support for typical for postgraduate study processes (e.g., assistance in literature search and expert identification);
- Organizational support (e.g., communication between researchers and experts, planning of scientific events).

Today, we can consider intelligent chat-bots as a crucial instrument of the digital transformation of science and education. However, their application in scientific institutions raises several challenges, some of which are specific to research activities:

- Responses with lacking clear argumentation and precisely defined references create risks of “hallucinations” and can lead to the use of incorrect background in research planning.
- Complex scientific terminology in queries can produce contradictory answers that use terms with semantics of other research domains.
- The use of open-access LLMs can violate the confidentiality and authorship of scientific data (e.g., caused by identifying reviewers for unpublished manuscripts).
- Existing information systems in research institutions are often based on outdated technologies, making integration with chat-bots difficult.

An available approach to addressing these challenges is the use of formalized knowledge of the research domain, represented through ontological models or semantically marked documents.

2. Literature Review

In current scientific literature, the terms “intelligent assistant” and “chat-bot” have acquired multiple meanings and require clear conceptual delineation. The absence of established definitions complicates their use in academic research and practical implementations. Therefore, it is important to examine how different authors define these concepts, what functional and methodological emphases they place, and which issues remain open for further investigation.

Analysis of contemporary publications demonstrates that the concept of an intelligent assistant is broader than chat-bot, because it encompasses knowledge management, analytics, and explainability, whereas a chat-bot is primarily a dialog interface that facilitates communication. The distinction between them is conceptual: a chat-bot is an interaction tool, while an intelligent assistant is a decision-support system.

An intelligent assistant is defined as a software system that integrates analytical models, knowledge bases, and interaction interfaces, providing users with relevant and explainable information. For example, in [3], the intelligent assistant is described as an AI-based system supporting personalized and adaptive learning.

Chat-bots, as a specific form of intelligent assistants, enable communication between users and information infrastructures. They facilitate automation of research project management, communication among their participants, and access to relevant databases and repositories [4]. Moreover, chat-bots can act as digital coordinators, reducing administrative workload and increasing the efficiency of collaborative work.

The main differences between the terms “chat-bot” and “intelligent assistant” are as follows:

- Level of knowledge integration: chat-bots provide access to information, whereas intelligent assistants use and integrate this information according to user needs (typically employing external knowledge sources such as ontologies, knowledge bases, and analytical models).
- Functional orientation: chat-bots are focused on communication, while intelligent assistants are oriented toward decision support and knowledge management.
- Presence of explanations: intelligent assistants provide clarification regarding the rules and sources by which responses are generated, whereas simpler chat-bots do not offer such explanations.

Thus, in contemporary scientific literature, the concepts of “intelligent assistant” and “chat-bot” are often used as interrelated, yet they emphasize different conceptual aspects. Chat-bots are primarily considered dialog interfaces that enable communication between the user and the information system, while intelligent assistants are defined as complex support systems integrated with knowledge bases, analytical models, and mechanisms of explainability. In addition to the user interface, their essential component is the functionality of the system itself.

Many researchers emphasize that in complex and dynamic domains, intelligent chat-bots require formalized representations of domain knowledge (e.g., ontological models, thesauri, taxonomies) with specific means of knowledge representation. Use of intelligent chat-bots requires to harmonize domain-specific terms and concepts [5], since without semantic models such chat-bots cannot maintain the relevance and reliability of recommendations in dynamic scientific environments [6].

Recent studies show that LLM-based chat-bots can provide more convenient user communication and adaptive responses [7]. However, their application is associated with several challenges:

- Explainability problem: LLMs often generate responses without clear reasoning that complicating their use in critical domains with complex knowledge structure and big volume of specific rules;
- Semantic consistency problem: the absence of domain-specific knowledge leads to contradictory, ambiguous and incorrect recommendations or to generation of trivial statements;
- Reliability problem: the risk of model “hallucinations,” when LLM produces information without real sources and even claims to use real sources (but verification shows that such sources do not exist or contain different and non-relevant information);
- Institutional integration problem: the difficulty of combining chat-bots with existing information systems and regulatory frameworks that are based on out-of-date information technologies.

These challenges highlight the necessity of transitioning to intelligent chat-bots that can employ formalized domain knowledge models, defining core concepts and relationships within the domain.

The development and implementation of a chat-bot can vary significantly depending on the technological stack and the complexity of the application domain. AI-based chat-bots are designed for natural language communication similar to human interaction. NLP methods are employed in chat-bots to understand user input by decomposing data into smaller fragments and to generate natural language responses [8]. The construction of chat-bots requires research in such areas of NL understanding and generation, as entity recognition (NER), and intent extraction (IE).

There are three NLP models commonly used in chat-bots:

- pattern-based,
- retrieval-based,
- deep-learning-based [9].

In pattern-based models, the chat-bot generates a response using a “question–answer” template that corresponds to the most relevant answer to the user’s query. However, such models are limited by their

functionality, because they cannot generate a response if the required templates or questions are absent from the database. Such restrictions significantly narrow the scope of their use

Retrieval-based models operate on principles of selecting responses from a set of predefined options using various methods, such as simple rule-based systems or ensemble machine learning techniques, to provide the most appropriate answer considering the context [10]. Although training these models on large datasets can reduce errors, they remain constrained by the fact that they can only generate responses from available options rather than produce new NL text. Moreover, their responses often lack naturalness, making them unsuitable for chat-bots that must handle complex queries.

Deep-learning-based models are increasingly applied in chat-bot development. One of the most widely studied models is the “sequence-to-sequence” (Seq2Seq) model that is a subclass of recurrent neural networks [11]. Originally designed for machine translation, Seq2Seq is now frequently used in conversational modeling. Its key advantage lies in the ability of the learning algorithm to transform input data into hidden vectors, which are then used to generate outputs. Another popular model is “Long Short-Term Memory” (LSTM), with a deep-learning algorithm based on recurrent neural networks, applicable to sequential data such as speech and video [12]. LSTM models are widely used in speech recognition; however, due to their computational complexity, they are unsuitable for domains requiring detailed explanation of prediction results. The Hybrid Emotion Interference Model (HEIM) represents another advanced NLP algorithm, designed to interpret human emotions, facial expressions, and voice signals. It employs deep, sparse neural networks to analyze diverse data patterns, including acoustic, textual, and geo-social data.

The technological foundations of intelligent assistants and chat-bots are formed at the intersection of several directions of scientific research:

- architectural models of dialog systems,
- large language models for recognizing of NL questions, deep processing of NL sources and generating of NL answers,
- semantic technologies for domain content structuring and retrieval of pertinent information resources,
- cloud platforms for scalability.

In scientific literature, several basic chat-bot architectures are distinguished:

- Rule-based systems, which rely on strictly defined rules and scenarios. They provide predictability and explainability but are limited in flexibility.
- Flow-based systems, which employ a tree-like structure where each node corresponds to a specific user intent. This approach is typically applied to handle standard queries, as it reduces resource consumption and minimizes database requests.
- Playbook-based systems, which are suitable for unpredictable dialogues where users may go beyond predefined topics. This approach is used for complex queries requiring integration with large language models.

Hybrid architectures allow combining the advantages of different approaches.

3. Problem Statement

The rapid expansion of intelligent assistants based on chat-bots in scientific and educational domains has created both opportunities and challenges. While traditional chat-bots primarily serve as dialog interfaces, intelligent assistants are expected to integrate analytical models, knowledge bases, and semantic technologies to provide explainable, personalized, and reliable support. In postgraduate education, the need for such systems is particularly acute: postgraduate students need in complex regulatory frameworks, diverse learning resources, and dynamic research environments. Existing solutions often fail to address issues of semantic consistency, explainability, and confidentiality, especially when relying solely on LLMs.

The problem addressed in this study is the design and implementation of an intelligent assistant that can overcome these limitations by combining LLMs with semantic technologies and ontological models of research domain. The task involves developing a hybrid architecture capable of handling predictable queries through flow-based agents and unpredictable dialogues through playbook-based agents, while ensuring transparency and trust via human-in-the-loop mechanisms.

4. Intelligent Assistant for Postgraduate Education

Researchers of the Institute of Software Systems of the National Academy of Sciences of Ukraine (ISS NASU) develop an intelligent assistant aimed to support the digitalization of doctoral and postgraduate education and to facilitate navigation through NL documents [13], including both regulatory and informational materials (https://iss.nas.gov.ua/aspirantura/abiturujentam_2025/pravila-priomu_2025). This assistant is integrated into the ecosystem of information support for postgraduate education (Fig. 1) considered in [2, 14] and uses a terminological system formalized in the ontological model of this ecosystem (concepts and relations between them). It is anticipated that this assistant can be scaled to serve the entire structure of the National Academy of Sciences of Ukraine, which coordinates the work of dozens of research institutes from different domains, ensuring the relevant task of integration of diverse science-support digital services.

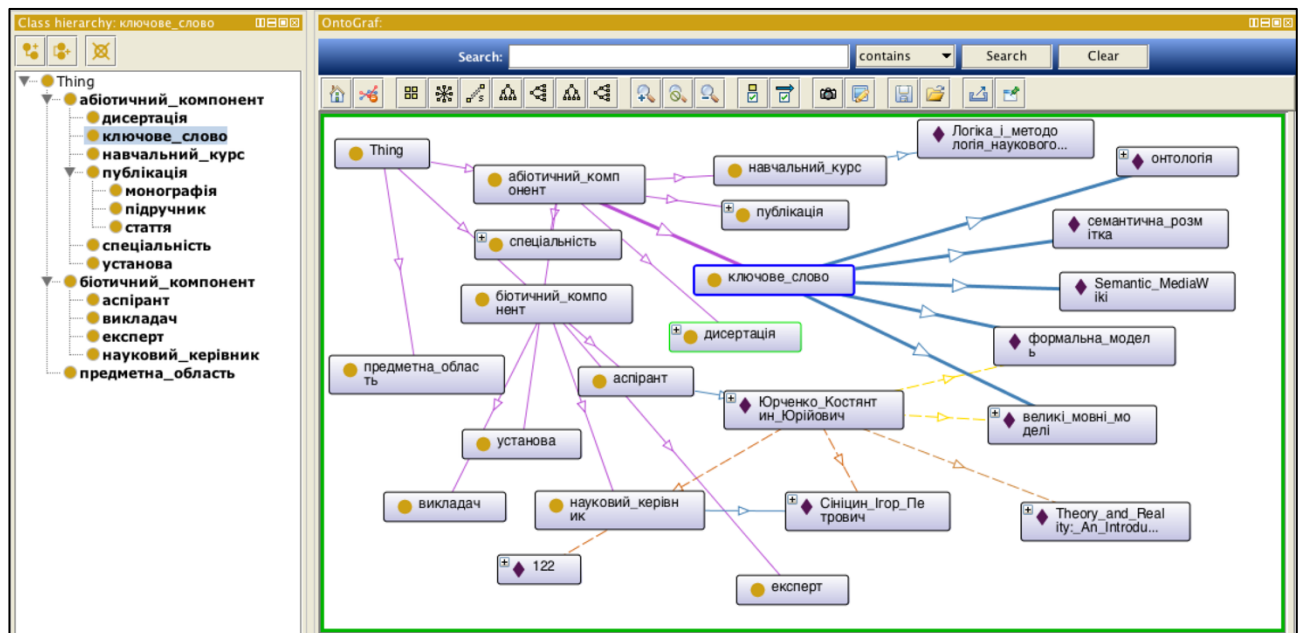


Figure 1: Ontological model of information support for postgraduate education of NASU

For the development of the intelligent assistant, the Vertex AI Agent Builder platform by Google was employed (<https://docs.cloud.google.com/vertex-ai/docs/start/introduction-unified-platform>). Owing to its extensive configuration options, this platform allows extending and branching the initial chat-bot structure, enabling the incorporation of new tasks or adjustments to existing ones based on accumulated practical experience. Importantly, modifications do not require changes to the initial settings of the assistant, thereby avoiding additional costs for reprogramming.

Vertex AI Agent Builder incorporates two main agent concepts used in building intelligent assistants:

- flow-based agents (<https://docs.cloud.google.com/dialogflow/cx/docs/quick/build-agent>);
- playbook-based agents (<https://docs.cloud.google.com/dialogflow/cx/docs/quick/build-agent-playbook>).

These agent concepts can utilize different types of LLMs.

A flow-based agent has a clear hierarchical structure (Fig. 2), represented as a continuous tree-like system where each link is connected to several subsequent links. Such agents are suitable for creating simple

conversational bots, where dialogue is maintained within one or several specific topics (intents), and even for collecting data (entities) from chat participants for further use within predefined topics.

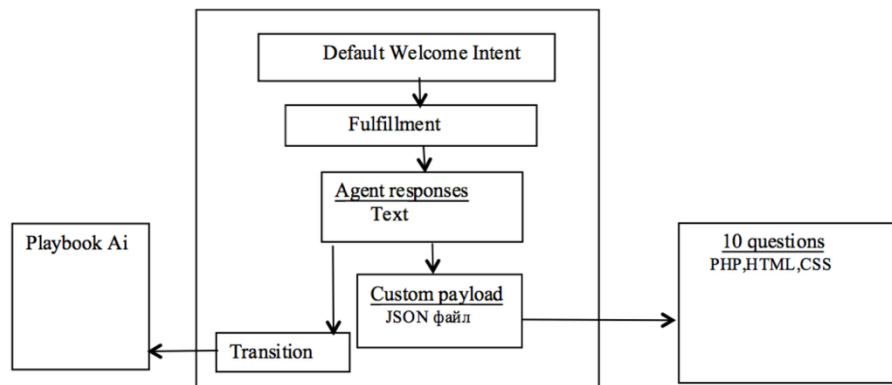


Figure 2: Block diagram of a Flow-based agent

A playbook-based agent is advisable in cases where the user’s dialogue with the intelligent assistant is less predictable, or when the assistant must provide responses based on a larger corpus of information and the trajectory of the conversation cannot be anticipated in advance (Fig. 3). Under this scheme, deviations from the main topic of discussion do not cause the assistant to “freeze,” and the dialogue appears more similar to human interaction.

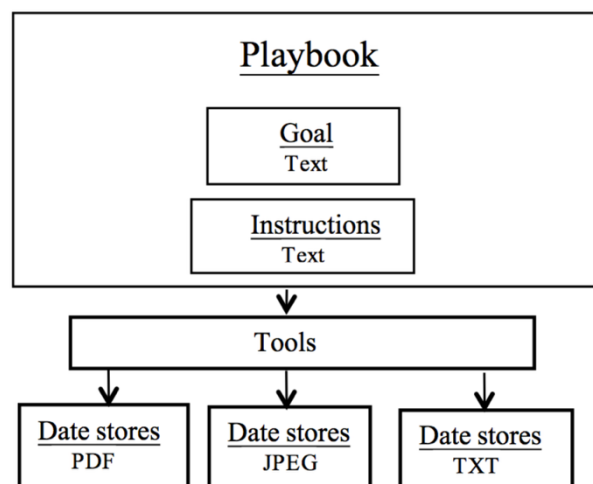


Figure 3: Block diagram of a Playbook-based agent

In the development of the intelligent assistant for supporting postgraduate education, we select a combined scheme that integrates both concepts (<https://docs.cloud.google.com/dialogflow/cx/docs/concept/data-store/hybrid>). This intelligent assistant is characterized by a set of architectural and functional features implemented within a multi-level hybrid architecture and enhanced by the capabilities of LLMs.

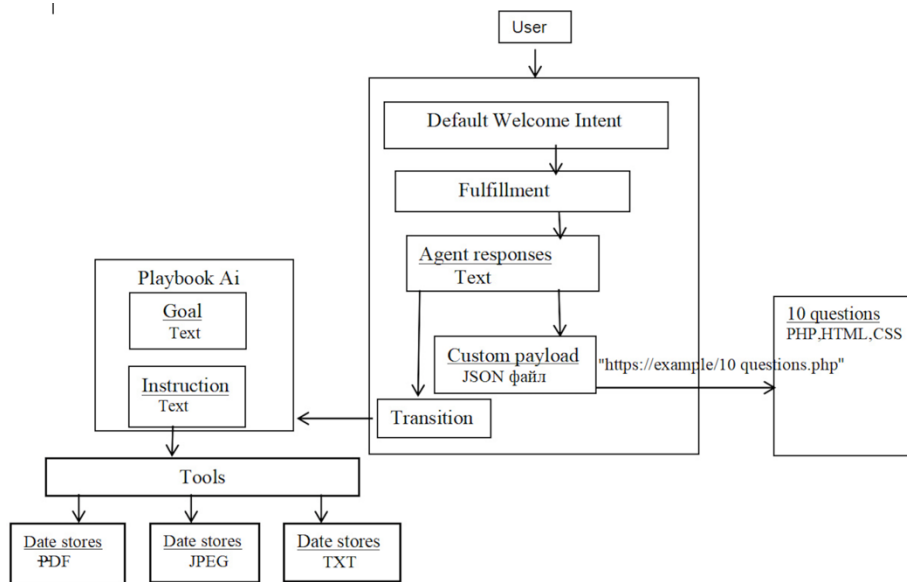


Figure 4: Block diagram of hybrid agent Flow-based+Playbook-based

The algorithm of the intelligent assistant operation is as follows:

- If the user is interested in specific questions (for example, regarding the conditions and general requirements of the National Academy of Sciences of Ukraine for graduate school), where the information is regulated by official NASU documents and the topic of conversation is clearly defined, the flow-based agent is activated.
- If the user is not satisfied with the list of typical questions, the dialogue is transferred under the control of the playbook-based agent.

This scheme is chosen from the point of view of resource efficiency. Since the intelligent assistant relies on a cloud-based data storage system, responding to each user query would otherwise require sending a request to the database and retrieving the corresponding answer, leading to additional computational and financial costs. Instead, the flow-based agent initially offers the user a link to a page with frequently asked questions hosted on the website, which does not require a database query. Subsequently, based on real query statistics, the list of typical questions can be expanded and refined.

The developed intelligent assistant is constructed as part of a comprehensive multi-level agent system. It is equipped with both internal and external interfaces. The internal interface includes a control panel for selecting LLM models, configuring prompts, setting objectives, and connecting data in various formats.

A key element of this software realization is the integration with the Gemini model that is used as the core for generating responses demonstrates its effectiveness in processing and synthesizing information, thereby expanding the functional capabilities of the system. Thus, the proposed environment contributes to increased efficiency in system development.

The external interface of the intelligent assistant, designed as a separate module, allows embedding the dialogue window both on web pages and in social networks. The pop-up window displays a field for entering user queries and a window for representation generated responses. An additional element of this interface is the interaction panel through which communication with the agent is carried out.

This user interface, implemented in the format of a dialogue window, ensures two-way communication, where the user can initiate a query and the agent provides a response. The system supports multilingual interaction, as demonstrated by dialogues conducted in both Ukrainian and English. Within this interface, a real conversation between the user and the intelligent assistant is illustrated (Fig. 5).

```

PHP JS HTML CSS

<link rel="stylesheet" href="https://www.gstatic.com/dialogflow-console/fast/df-messenger/prod/v1/themes/df-messenger-
default.css">
<script src="https://www.gstatic.com/dialogflow-console/fast/df-messenger/prod/v1/df-messenger.js"></script>
<df-messenger
  project-id="sound-fastness-427809-q8"
  agent-id="f42bdce5-bdf2-4e59-8170-d4325a56bd7f"
  language-code="uk"
  max-query-length="1">
  <df-messenger-chat-bubble
    chat-title="AI-ASISTEENT ISS NASU">
  </df-messenger-chat-bubble>
</df-messenger>
<style>
df-messenger {
  z-index: 999;
  position: fixed;
  --df-messenger-font-color: #000;
  --df-messenger-font-family: Google Sans;
  --df-messenger-chat-background: #9FEED7;
  --df-messenger-message-user-background: #F66B6B;
  --df-messenger-message-bot-background: #15CDD9;
  --df-messenger-titlebar-title-font-size: 18px;
  --df-messenger-titlebar-background:#9A9595;
  --df-messenger-chat-padding: 10px;
  bottom: 1px;
  right: 16px;
}
</style>

```

Figure 5: External interface of the intelligent assistant (code)

The development of a chat-bot for doctoral student support is envisaged as a pilot implementation within the broader project entitled “Intelligent Personal Assistant for Researchers” (IPAR) that is based on the technological platform Linza. This concept represents a novel approach to the automation and personalization of scientific workflows in the context of the National Academy of Sciences of Ukraine. We consider IPAR not as a generic conversational interface, but as a structured, ontology-based environment that integrates LLMs, semantic technologies, and modular architectures in order to provide comprehensive support for routine academic tasks. User ontologies represent research interests and methodological preferences of individual researcher, domain ontologies represent the conceptual structures of specific scientific fields, and task ontologies model recurrent academic processes such as literature review, reporting, or peer feedback, and contextual ontologies that reflect situational factors such as institutional regulations or collaborative environments. The perspectives of this research are therefore directed towards the creation of a sustainable framework where researchers and postdoctoral students can rely on intelligent systems to optimize their engagement with scientific resources, while simultaneously ensuring compliance with institutional requirements and safeguarding the integrity of their scholarly work.

5. Place of Intelligent Chat-bots in Complex Experience Extraction Platform Linza

The Linza project described in details in [14] uses hybrid chat-bots based on the large language models in a real-world environment. Its experience allows not only to confirm the effectiveness of the proposed approaches, but also to identify a number of limitations and research questions that require further development.

Linza is developed as an instrument oriented on complex *Experience Extraction from Documents (EED)* tasks. In this research we define as a main criterion that unites different applied tasks of EED its orientation toward the result where methods of information analysis and procedures for processing input information are not formalized and largely depend on the non-formal user’s requirements for the resulting object described by examples and their evaluations. These user requirements can be fixed by NL that allow for ambiguous interpretation and require NLP means for creation of formal rules.

The technological platform Linza has to perform EED tasks of varying complexity. Its main goal is the intelligent processing of heterogeneous unstructured documents to transform them into resulting complex information objects (CIOs) in accordance with a broad range of user requirements [14]. This functionality is achieved by combining semantic technologies, which allow the system to understand the meaning of information, and LLMs provide linguistic analysis of document content and generation of natural language texts. The main purpose of this system is to transform raw NL data into the integrated knowledge base that supports the generation of clear, structured, and comprehensible reports and recommendations.

Linza applies the human-in-the-loop concept, ensuring full transparency and control. A human expert can monitor and validate each stage of processing and the obtained results, with access to all necessary information, but without visibility into other users' documents. This feature not only increases trust in automatically generated documents but also guarantees factual accuracy, which is essential for responsible decision-making.

The Linza system employs a combination of Semantic MediaWiki (SMW) technologies and means of knowledge formalization based on ontological models that are used as a base for semantic markup, together with generative artificial intelligence tools for analyzing NL documents and acquiring from these documents the values of the semantic properties. Use of the wiki-based platform makes the knowledge acquired by various LLMs from documents visible, understandable, and editable. Domain ontology helps in control valid combinations of property values and identify those ones that do not meet domain constraints. Such integration combines the power of LLMs with the reliability and explainability of semantic analysis results.

Main functional capabilities that single out Linza from other EED means:

- Automated knowledge-based processing and structuring of raw data for routine tasks is combined by expert-based analysis of domain knowledge structure;
- recommendations for users are generated of base on samples, evaluated examples and additional information from user profiles;
- system propose explanations for these user recommendations that support multi-level control (unlike the “black box” approach of LLMs) where user can analyze intermediate results of data interpretation based on semantic annotation of NL data;
- dynamic adaptability to domain changes and user requirements, analysis of modifications of user needs.

Now Linza functionality is represented by such set of functional modules:

- Module that generates and updates the task domain thesaurus – the set of main concepts and relations used for representation of user requirements and describing of required result;
- repository module that stores input and output NL documents on various stages of their structuring and their task-specific metadata;
- Module that generates semantic annotation of NL documents and enriches them by semantic markup (based on task thesaurus) for further analysis, navigation and retrieval;
- SMW resource that contains current versions of semantically marked NL documents where thesaurus concepts are used as names of semantic properties;
- Module that generates rules for result constructing (on base of user requirements formalization or by traductive generalization of evaluated examples));
- Module that generates explanations of results (as a sequence of used rules for result constructing and selected data resources);
- Intelligent chat-bot that supports user NL interaction with system.

The module of domain thesaurus generation provides knowledge-based instruments for constructing a consistent dictionary that contains definitions of domain concepts and some relations among them. This representation of task knowledge is based on the ontological model of this task domain, but it is not including axioms and structures definitions of concepts that complicate data processing on this stage. Thesaurus helps all participants to interpret data semantics unambiguously. Depending on the task specific and presence of appropriate domain ontology, the thesaurus can be generated directly from ontology as a subset of its classes and instances or is formed through NL document analysis with the help of LLMs or other NL processing tools. If necessary, it can be edited manually by users or domain experts.

The repository module ensures the storage of documents at all stages of their processing and analysis. Document annotations contain all necessary information for creation SMW markup, but repository supports more wide set of functions that can be used to investigate the origin of data, its quality, issues of authorship and rights of use by different groups of users. Such a repository supports access to the initial version of documents, which provides the ability to check the quality of recognition of semantic properties and allows processing of documents according to modified requirements (for example, with a different thesaurus or for a different problem statement).

The semantic markup module transforms natural language documents into structured representations by linking content elements to thesaurus concepts. The results of this processing are represented in a form `[[thesaurus_concept:: text-based value| NL text fragment]]`. LLMs are applied to identify relevant fragments. Structured documents can then be placed into the Semantic Wiki resource that supports user-oriented instruments of search, navigation, and data integration capabilities. This simplifies expert operations for evaluating annotation quality and improving LLM performance.

The SMW resource contains documents with semantic markup. SMW resource contain documents that are represented as separate Wiki pages with semantic markup: names of semantic properties are selected from task thesaurus, and their values are acquired from document content. Unlike a repository module of Linza, only one version of the document is stored in SMW, and the form of presentation of structural elements is limited by the expressive capabilities of the SMW. But the advantage of using the SMW is an intuitive user interface, similar to the widely known Wikipedia, built-in knowledge export tools and the ability to collaboratively supplement content with other means. Thus, a wiki resource can be considered as one of the final results of the work of the Linza, which is focused on a specific user task and which can be changed using other modules in cases where the information environment or user needs change.

Module that generates explanations of results proposed to user the sequence of rules from the set that is created for current task by module of rule generation on base of examples. This sequence shows how input data were transformed into result object and what data resources were used as inputs for these rules.

In addition, Linza provides an operational environment that is developed to coordinate all stages of document processing – from collection and analysis to transformation and verification of data. It has to support secure and reliable data exchange among system modules responsible for uploading and preprocessing documents, semantic annotation, knowledge modeling, response generation using LLMs, and information integration. Another important function of this environment is access management for authorized users. The system protects data through role-based access, complete logging of all actions, and secure HTTPS connections, preventing unauthorized modifications or deletions of data.

Currently, the Linza technology is in the development stage. The functionality for the main modules has been fully defined; technological means and data formats have been selected. Research prototypes of these modules have been created and tested on various data sets. It is important that for most modules several different implementation options are envisaged, depending on the complexity and specific requirements of individual practical tasks. However, the task of automated data exchange between modules is still under research.

From the point of view of the intelligent chat-bot development we can define such main aspects of other Linza modules:

- Domain thesaurus generation module proposes the set of task concepts and relations that chat-bot has to apply in communication with user;

- Document storage module provides semi-structured NL documents that can be enriched by task-specific semantic properties for SMW resource;
- Document semantic markup module automatically produces this additional semantic markup according to domain-specific rules and syntax of SMW;
- SMW resource is a source of user-specific (according to information from user profile and current task) documents that can be generated automatically by SMW queries;
- Rule generation module provides elements of task knowledge for chat-bot functioning and for LLM prompts;
- Result explanation module provides examples of sequence that contain rules for result constructing and selected data resources that chat-bot can propose to user in more understandable manner.

The Linza project can be understood as a general technology for acquiring experience through the integration of semantic wikis with LLMs. Its conceptual novelty lies in the fact that it does not merely provide a framework for combining formalized knowledge and statistical models, but also generates concrete outcomes, one of which is the development of intelligent chat-bots. These, chat-bots are not the ultimate goal but rather a derivative product of a broader ecosystem shaped by Linza that help users in understanding and application of knowledge generated by Linza.

Linza demonstrates the feasibility of systematical combining semantic knowledge bases, represented in the format of SMW, with the generative capabilities of LLMs. This integration enables not only the provision of user responses but also the construction of explanations for how these responses are generated, and chat-bots can function as tools that clarify the sources and logic behind recommendations.

Modular nature of Linza opens perspectives for applying other intelligent technologies. For instance, the integration of knowledge graphs constructed on base of academic publications could significantly enhance the accuracy and relevance of chat-bot responses in sphere of postgraduate education. Other interesting direction of Linza development concerns integration with Knowledge-Infused Learning (K-IL) [15]. The central idea of K-IL is that deep learning has to be complemented by external knowledge – ontologies, knowledge graphs, and semantic structures – to overcome the limitations of purely statistical generalization, and we can export these knowledge structures from SMW resource and by processing of document metadata from repository. This approach implemented in Linza enables the intelligent chat-bot to propose personalized educational trajectories for postgraduate students. These trajectories are constructed from a repository of semantically annotated documents and transformation rules, enriched with metadata.

Use of Linza technology underscores the necessity of modeling the postgraduate education ecosystem with explicit consideration of chat-bots. Within this ecosystem, chat-bots are not isolated tools but integrated agents that interact with regulatory documents, curricula, databases, and communication platforms. Modeling such an ecosystem has to account for the information lifecycle of various documents.

Therefore, we can consider Linza as a platform for further research that demonstrates how intelligent chat-bots can enrich possibilities of this technological ecosystem to supports postgraduate education processes.

6. Perspectives for Further Research

We analyze the potential shortcomings of the proposed solution, possible directions for integrating with other intelligent technologies, and a substantiation of the need to model the postgraduate education ecosystem with explicit consideration of chat-bots.

Proposed solution has some potential technical and methodological shortcomings:

- The architecture that relies on a combination of flow-based and playbook-based agents with LLMs as the core is vulnerable to the phenomenon of model “hallucinations,” manifested as the generation of false or unverified information. This vulnerability necessitates the development of formalized mechanisms for verification of responses, the integration of sources with trusted attestation, and the construction of quantitative trust metrics that would allow assessing the reliability of each response.

- Dependence on commercial LLM platforms introduces risks related to service availability, changes in usage policies, and scaling costs. These considerations justify research of hybrid approaches that combine lightweight local models for routine operations with remote high-capacity models for complex queries, as well as economic evaluations of such configurations.
- Integration with existing institutional information systems is complicated by the absence of unified data exchange standards and semantic interoperability, which underscores the need to develop universal APIs, semantic layers and metadata profiles tailored to the specificities of the scientific and educational domain.
- Issues of confidentiality and protection of personal and research data remain critical: the use of open LLMs can lead to leakage of sensitive information, and therefore research is required on privacy-preserving methods (for example, federated learning and differential privacy) and on the formulation of access and audit policies that comply with academic ethics and legal requirements.

We consider opportunities and possible advantages for future applying other intelligent technologies (such as knowledge graphs, ontologies, semantic markup, retrieval-augmented generation, machine learning etc.) for intelligent assistant development.

The integration of knowledge in the form of knowledge graphs and ontologies can increase semantic consistency of responses and provide traceability of sources: research should focus on methods for automatic construction and updating of such graphs from scientific publications from selected scientific domain, regulatory documents and institutional databases.

Neuro-symbolic approaches can combine the flexibility of LLMs with formal logic and rule-based reasoning, thereby enhancing explainability and enabling formalized verification procedures, which is particularly important in contexts that require legally or ethically grounded decisions.

The application of information extraction and retrieval-augmented generation (RAG) techniques with indexing of verified sources can reduce the risk of producing unreliable answers; research tasks include optimizing document sampling strategies, source ranking algorithms and the integration of citation mechanisms into generated responses.

The implementation of modules for affective state analysis and adaptive interaction (emotion-aware dialogue systems) can improve the quality of user support in stressful or conflictual situations, but such developments require rigorous ethical assessment and safeguards against potential manipulation. Moreover, in future we plan to consider various aspects of socio-psycho-physiological study of personality that can make dialog more personified.

The necessity of modeling the postgraduate education ecosystem with consideration of chat-bots is aimed on formal definition of role and tasks of intelligent assistants in this process and is caused by need in more formal definitions of recommendation representation. Effective deployment of chat-bots in postgraduate education requires systemic modeling of the ecosystem that encompasses not only technical components but also organizational processes, stakeholder roles (postgraduate students, scientific advisers, administration, lecturers, external and internal experts, etc.), information flows and data access regulations.

Absence of such modeling causes risks of functional duplication, access conflicts and process inconsistencies will remain high. moreover, such modeling should include information lifecycle scenarios – from the creation of a scholarly artifact through its storage, review and publication – with identification of interaction points where a chat-bot can deliver the greatest added value (for example, preliminary reviewer selection, automated compliance checks with formal requirements, and personalized learning trajectories).

Use of the formal model of this ecosystem can provide mechanisms for human supervision (“human-in-the-loop”) in critical processes to ensure accountability and the possibility of correcting automatically generated decisions. This research should determine optimal policies for allocating responsibilities between automated agents and human experts in planning of personal learning trajectories for postgraduate students and define the place of intelligent assistants in it.

This analysis becomes the base for defining of specific directions for further research and experimental work. We plan to test the reliability of chat-bot responses in scientific contexts with test cases that simulate queries from diverse disciplines (based on existing sets of documents from multiple institutions) with different levels of request complexity. More attention would be paid to methods for integration of domain

ontologies and knowledge graphs with LLMs, with emphasis on automatic knowledge updating and terminology alignment and to empirical research on the pedagogical effectiveness of personalized learning trajectories generated by chat-bots for postgraduate students.

In summary, future research has to be interdisciplinary, combining technical innovation with organizational, ethical and pedagogical approaches, and should aim at the creation of open standards and tools that ensure interoperability and resilience of solutions. Proposed systemic approach to modeling the postgraduate education ecosystem – one that accounts for the technical limitations of LLMs, user needs, confidentiality requirements and economic feasibility – will enable the transformation of chat-bots from experimental instruments into reliable components of academic infrastructure.

7. Conclusions

The development and implementation of intelligent chat-bots in scientific activities and postgraduate education demonstrate substantial potential for reshaping academic infrastructures. This study confirms that the integration of large language models (LLMs) with semantic technologies provides a new quality of information support that surpasses traditional dialog systems. Intelligent assistants not only facilitate access to information but also generate explainable recommendations, integrate knowledge from diverse sources, and support personalized educational trajectories.

The analysis of scientific publications in the field of developing and applying intelligent assistants in scientific research activities highlights the relevance of this study and the appropriateness of employing large language models to expand the functionality of such systems. The experience of implementing an intelligent assistant for postgraduate education support at the Institute of Software Systems of the National Academy of Sciences of Ukraine, along with its practical use, substantiates the feasibility of scaling and replicating this approach in more complex projects – for example, in maintaining the information infrastructure of the National Academy of Sciences of Ukraine. The proposed software solution is aimed at effectively meeting the informational needs of users while optimizing computational resource consumption.

CRedit Author Statement

IS: Conceptualization, Methodology;

PA: Supervision, Methodology;

JR: Formal analysis, Methodology, Writing – Original draft preparation;

BL: Software, Data curation, Visualization,

OS: Validation.

Declaration on Generative AI

During the preparation of this work, the authors used AI programs Chat GPT 4o and Copilot for grammar and spelling check. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] L. Labadze, M. Grigolia, L. Machaidze, Role of AI chat-bots in education: systematic literature review. In: International Journal of Educational Technology in Higher Education, 20(56), (2023). doi: 10.1186/s41239-023-00456-7
- [2] I.P. Sinitsyn, J.V. Rogushina, K.U. Urchenko, Integration of large language models with semantic processing tools as an instrument for knowledge digitization. In: Problems in Programming, 2, (2025), 63–76. URL: <https://pp.isoftware.kiev.ua/index.php/ojs1/article/download/838/889>. (in Ukrainian).
- [3] R. Sajja, Y. Sermet, M. Cikmaz, D. Cwierny, I. Demir, I., Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. In: J. Information, 15(10), (2024). doi: 10.3390/info15100596.

- [4] S. Ceccacci, C. Giaconi, N. Del Bianco, AI Chat-bots in Education: Challenges and Opportunities. In: *J. Information*, 16(3), (2025), 235. MDPI. doi: 10.3390/info16030235.
- [5] Y. Zhang, H. Li, X. Chen, Large-Language-Model-Enabled Text Semantic Communication Systems. In: *Entropy*, 26(5), (2024), 512. MDPI. doi: 10.3390/e26050512.
- [6] R. Mohandoss, Context-based semantic caching for LLM applications. In: *2024 IEEE Conference on Artificial Intelligence (CAI)*, (2024), pp. 371-376. IEEE Press. doi: 10.1109/TKDE.2024.1234567.
- [7] S. K. Dam, C. S. Hong, Y. Qiao, C. Zhang, A Complete Survey on LLM-based AI Chat-bots. arXiv preprint arXiv:2406.16937. Cornell University. 2024. doi: 10.48550/arXiv.2406.16937.
- [8] A. Eltahir, H. Abdulla, J. Platos, V. Snasel, An overview of Chat-bot structure and source algorithms. In: *2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, (2022), pp. 91-98.
- [9] H. T. Hien, P. N. Cuong, L. N. Nam, H. L. Nhung, L. D. Thang, Intelligent assistants in higher-education environments: the FIT-EBot, a chat-bot for administrative and learning support. In: *Proceedings of the 9th international symposium on information and communication technology*, 2018, pp.69-76. doi: 10.1145/3287921.32879.
- [10] S. Pandey, S. Sharma, A comparative study of retrieval-based and generative-based chat-bots using Deep Learning and Machine Learning. In: *Healthcare Analytics*, 3, 100198 (2023). URL: <https://www.sciencedirect.com/science/article/pii/S2772442523000655>.
- [11] K. Palasundram, N. M. Sharef, N. Nasharuddin, K. Kasmiran, A. Azman, A.: Sequence to sequence model performance for education chat-bot. In: *International journal of emerging Technologies in Learning (iJET)*, 14(24), (2019), 56–68.
- [12] K. K. Nirala, N. K. Singh, V. S. Purani, A survey on providing customer and public administration-based services using AI: chat-bot. In: *J. Multimedia Tools and Applications*, 81(16), (2022), 22215-22246.
- [13] I. Sinitsyn, K. Yurchenko, B. Laschonov, Y. Bova, Architectural solution for a chat-bot with recursive knowledge refinement based on LLAMA 3 and Semantic MediaWiki. In: *Tavria Scientific Bulletin. Series: Technical Sciences*. 4, 2025, vol. 1. 270–278. doi: 10.32782/tnv-tech.2025.4.1.27 (in Ukrainian).
- [14] I. Sinitsyn, Rogushina, J., K. Yurchenko, Y. Bova, Integration of Semantic Wiki Technologies with Large Language Models as a Technological Foundation for Experience Acquisition from Natural Language Documents, 9th International Scientific and Practical Conference “Applied Information Systems and Technologies in the Digital Society” AISTDS'2025, Kyiv, Ukraine, October 1, 2025, CEUR Workshop Proceedings, 4133, 2025, pp.115-135, URL: ceur-ws.org/Vol-4133/S_10_Rogushina.pdf.
- [15] U. Kursuncu, M. Gaur, A. Sheth, Knowledge infused learning (K-IL): Towards deep incorporation of knowledge in deep learning. (2019). URL: <https://arxiv.org/pdf/1912.00512>.