DOI: https://doi.org/10.36910/6775-2524-0560-2026-63-08
UDC 004.89:004.932
Datsok Yevheniia, Bachelor's Student
https://orcid.org/0009-0008-5101-5217
Yakovleva Olena, Ph.D., Associate Professor
https://orcid.org/0000-0002-6129-6146
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

GRADIO FOR RAPID PROTOTYPING OF MULTIMODAL AI SYSTEMS ON THE EXAMPLE OF MEDICINAL PRODUCT PACKAGING INFORMATION RECOGNITION

Datsok Y., Yakovleva O. Gradio for Rapid Prototyping of Multimodal AI Systems on the Example of Medicinal Product Packaging Information Recognition. This paper investigates the organization of a rapid prototyping environment for multimodal medicinal product recognition systems based on cloud MLLM services and the Gradio library. The study addresses the complexity of developing experimental AI systems associated with frontend and backend implementation, API integration, and model testing workflows. A research-oriented environment is proposed in which Gradio acts as an intermediate interaction layer between the user and a multimodal language model. The developed environment enables rapid creation of MVP prototypes for recognizing information from photographs of medicinal product packaging, supports multimodal data processing, and allows interactive testing of different model configurations without modifying the system architecture. Particular attention is paid to automatic dataset accumulation during user-system interaction. The main practical results include a lightweight architecture for multimodal AI prototyping, integration of Gradio with cloud MLLM services, and support for interactive testing of prompts and models. The obtained findings demonstrate that Gradio reduces implementation complexity, accelerates development iterations, and simplifies experimentation with multimodal AI systems in medicinal product recognition tasks.
Keywords: multimodal language models, MLLM, Gradio, rapid prototyping, medicinal product recognition, multimodal systems, AI systems, image processing, cloud services, dataset accumulation.

Дацок Є.О., Яковлева О.В. Gradio для швидкого прототипування мультимодальних AI-систем на прикладі задачі розпізнавання інформації з упаковок лікарських засобів. У статті досліджено підхід до організації середовища швидкого прототипування мультимодальних систем розпізнавання лікарських засобів на основі хмарних MLLM-сервісів та бібліотеки Gradio. Розглянуто проблему складності створення експериментальних AI-систем, пов'язану з реалізацією frontend- та backend-компонентів, інтеграцією API та організацією тестування моделей. Запропоновано дослідницьке середовище, у якому Gradio використовується як проміжний шар взаємодії між користувачем і мультимодальною мовною моделлю. Розроблене середовище забезпечує швидке створення MVP-прототипів для розпізнавання інформації з фотографій упаковок лікарських засобів, підтримує обробку мультимодальних даних та дозволяє виконувати інтерактивне тестування різних конфігурацій моделей без модифікації архітектури системи. Окрему увагу приділено автоматичному накопиченню експериментальних даних у процесі взаємодії користувача із системою, що дозволяє поєднати rapid prototyping та формування експериментального датасету в межах єдиного дослідницького середовища. Основними результатами роботи є lightweight-архітектура мультимодального AI-середовища, інтеграція Gradio з хмарними MLLM-сервісами та підтримка інтерактивного тестування моделей і prompt-запитів. Отримані результати підтверджують доцільність використання Gradio для rapid prototyping мультимодальних AI-систем у задачах розпізнавання інформації з упаковок лікарських засобів.
Ключові слова: мультимодальні мовні моделі, MLLM, Gradio, швидке прототипування, розпізнавання лікарських засобів, мультимодальні системи, AI-системи, обробка зображень, хмарні сервіси, накопичення датасетів.

Problem statement. Modern multimodal large language models (MLLMs) demonstrate high performance in image recognition and information processing tasks, opening new opportunities for automated data input in applied information systems [1]. Similar approaches are already used in the financial domain, particularly in systems for automated receipt recognition and structured information extraction, such as Billka AI [2]. The use of multimodal models makes it possible to simultaneously process textual and visual information, significantly improving the efficiency of document and image analysis.

One of the promising areas of MLLM application is the recognition of information from photographs of medicinal product packaging, including medicine names, manufacturers, expiration dates, and other characteristics [10]. The use of such approaches reduces the amount of manual data entry, simplifies household medication management, and improves user interaction with information systems.

At the same time, the development of experimental multimodal AI systems remains a complex task. Traditional approaches to implementing such systems require separate frontend and backend components, cloud API integration, interaction management with AI models, and dedicated testing interfaces. This significantly increases the complexity of MVP prototype development, complicates experimentation, and slows down the process of testing different model configurations.

Particularly relevant is the problem of organizing a research environment that enables rapid creation and modification of multimodal systems, interactive model testing, and accumulation of experimental results without developing a full-scale web architecture. In this context, the use of lightweight solutions for rapid AI system prototyping is considered appropriate.

Analysis of research. Traditional OCR-based approaches are commonly used for automated information extraction from images, including Tesseract, which is an open-source OCR engine for text recognition from digital images and documents [3]. Such tools are effective for basic text extraction tasks; however, their application in medicinal product packaging analysis remains limited because they primarily operate on isolated text fragments and do not consider the complete visual context of the image.

The development of multimodal large language models has significantly expanded image processing capabilities, as MLLMs are able to combine textual and visual information analysis. Recent studies indicate that MLLMs demonstrate new capabilities in OCR-free reasoning, image understanding, and integration of visual context with language processing [4]. This makes them promising for medicinal product packaging recognition tasks, where not only individual text labels but also their spatial arrangement, visual context, and relationships between packaging elements are important.

Another important research direction involves the organization of rapid prototyping environments for AI systems. Various tools can be used for this purpose, including Gradio, Streamlit, Hugging Face Spaces, and traditional web-oriented approaches. Gradio is positioned as a Python library for rapidly creating web interfaces for machine learning models and APIs without implementing a separate frontend component [5]. Streamlit is also an open-source Python framework for creating interactive data applications with minimal code, making it suitable for analytical systems and dashboard-oriented solutions [6]. Hugging Face Spaces [7], in turn, is used as a platform for hosting models, datasets, and demonstration AI applications.

A comparison of approaches to organizing rapid prototyping environments for AI systems is presented in Table 1.

Table 1. Comparison of approaches to rapid prototyping of AI systems

Characteristic	Traditional Web Stack	Streamlit	Hugging Face Spaces	Gradio
Separate frontend required	yes	no	no	no
Python integration	indirect	direct	depends on implementation	direct
Orientation toward ML/AI models	medium	medium	high	high
Support for multimodal scenarios	depends on implementation	medium	high	high
MVP development speed	medium/low	high	high	high
Interactive model testing	limited	supported	supported	supported
Prompt engineering support	depends on implementation	medium	medium	high
Implementation complexity	high	low	medium	low
Suitability for MLLM research environments	medium	medium	high	high

The analysis shows that Gradio is not the only tool available for rapid AI system prototyping; however, it provides several advantages specifically for research-oriented multimodal scenarios. It is particularly suitable when there is a need to quickly create interfaces for image uploading, interaction with

Python-based processing logic, prompt testing, visualization of structured recognition results, and further accumulation of experimental data.

Therefore, the use of Gradio not only as a demonstration tool for ML models but also as a lightweight environment for rapid prototyping, interactive MLLM testing, and experimental dataset generation during user–system interaction remains insufficiently explored.

Purpose of the study. The purpose of this study is to investigate an approach to organizing a rapid prototyping environment for multimodal medicinal product recognition systems using the Gradio library and cloud-based MLLM services, as well as to evaluate the feasibility of applying such an approach for the development of research-oriented AI systems.

The main scientific and practical results of the study are as follows:

– a lightweight architecture for a rapid prototyping environment for multimodal AI systems has been proposed;

– integration of Gradio with cloud-based MLLM services for processing photographs of medicinal product packaging has been implemented;

– a mechanism for experimental dataset accumulation during user–system interaction has been developed;

– support for interactive testing of different model configurations and prompt queries has been provided.

Presentation of the main research material and justification of the obtained results.

The architecture of a rapid prototyping environment for multimodal medicinal product recognition systems should provide the ability to quickly create, test, and modify experimental AI solutions without developing a full-scale client–server infrastructure. Unlike the traditional approach, which requires separate implementation of frontend interfaces, backend services, API layers, and result storage modules, the proposed environment is focused on minimizing auxiliary components and concentrating on the investigation of multimodal model performance.

The proposed architecture is based on the Gradio library [11], which serves as an intermediate layer between the user and the data processing logic. Gradio [5, 11] provides a web interface for uploading photographs of medicinal product packaging, launching the recognition process, and displaying the obtained results in a form suitable for analysis. Such an approach enables rapid creation of MVP prototypes without implementing a separate frontend component.

The functional logic of the environment is implemented using Python [8]. It includes preprocessing of input images, generation of structured prompts for multimodal language models, transmission of requests to cloud-based MLLM services, processing of the obtained responses, and transformation of recognition results into a structured format. Within the proposed approach, the recognition result can be represented as a JSON object [9] containing the main medicinal product characteristics, including medicine name, dosage, manufacturer, expiration date, active substance, and other attributes available on the packaging.

The cloud-based MLLM service performs the role of the main intelligent component of the system. It receives a multimodal request consisting of an image and a textual instruction, analyzes visual and textual information from the packaging, and generates a response according to the predefined structure. The use of cloud services eliminates the need for local model deployment, reducing hardware requirements and simplifying experimental testing of different models.

Another important architectural component is the experimental dataset accumulation module. Its purpose is to store input images, recognition results, processing time, and additional metadata. Such a module transforms the rapid prototyping environment not only into a demonstration tool for model capabilities but also into a dataset generation mechanism for further analysis of recognition quality.

The overall structure of the proposed environment can be represented as a sequence of interaction between the following components: user, Gradio interface, Python processing module, cloud-based MLLM service, and experimental result storage. This architecture is sufficiently simple for rapid deployment while simultaneously providing the required functionality for conducting experiments, comparing models, and accumulating experimental results. The structural architecture diagram is presented in Fig. 1.

In the proposed rapid prototyping environment, interaction with multimodal large language models is implemented through APIs of cloud-based MLLM services. Such an approach makes it possible to use the capabilities of modern multimodal models without deploying complex local AI infrastructure and enables rapid testing of different model configurations within a unified research environment.

The interaction process is based on a multimodal request consisting of an image of medicinal product packaging and a textual instruction (prompt). The image is transmitted to the model in encoded form, after which the MLLM analyzes visual information and textual elements presented on the packaging. The textual instruction defines the structure and format of the model response, as well as the set of characteristics that must be extracted from the image.

Within the study, a structured approach to prompt engineering was applied. The model is instructed to return the recognition result in the form of a JSON object with predefined fields, which simplifies further processing of the obtained data. Depending on the specific task, the response structure may contain the medicine name, dosage, manufacturer, expiration date, active substance, and other available packaging attributes.

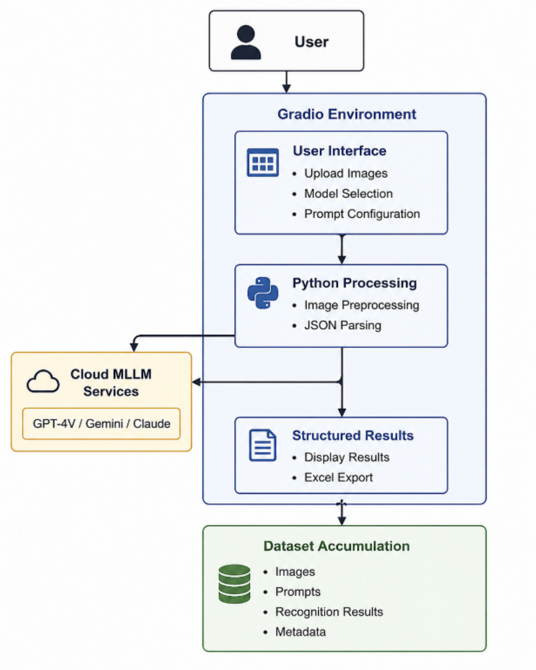

Fig. 1. Architecture of the rapid prototyping environment for a multimodal medicinal product recognition system

After receiving the response from the MLLM service, automatic processing is performed using Python tools. The resulting JSON object can be used for visualization within the Gradio web interface, storage in a local repository, or accumulation of experimental datasets. Such an approach standardizes recognition results and enables further comparison of different models and prompt configurations.

The use of APIs of cloud-based MLLM services also enables rapid testing of different multimodal models without modifying the system architecture. Different models can be used within the research environment depending on the task, performance requirements, or recognition quality. This provides flexibility for experimental workflows and simplifies evaluation of multimodal approaches in medicinal product recognition tasks. The sequence diagram of component interaction is presented in Fig. 2.

One of the key features of the proposed research environment is the use of the Gradio library as a tool for rapid creation of MVP prototypes of multimodal AI systems. Unlike the traditional web application development approach, which requires separate implementation of frontend and backend components, Gradio makes it possible to create an interactive web environment directly using Python. This significantly reduces the complexity of implementing experimental systems and shortens development time.

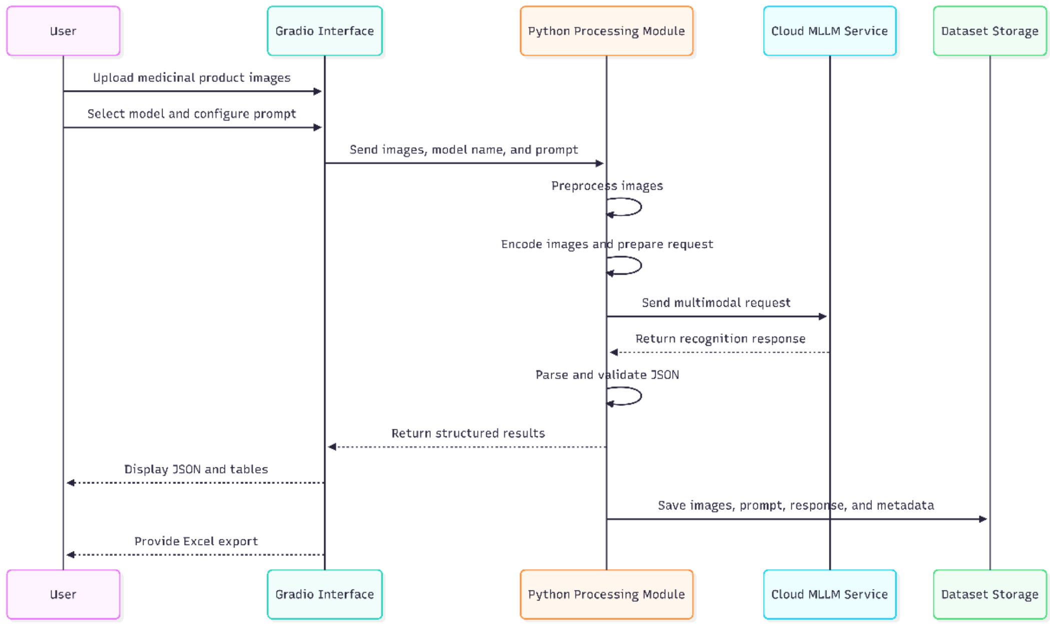

Fig. 2. Sequence of interaction between the user, Gradio, and the MLLM service

Within the study, a prototype of a multimodal medicinal product recognition system was implemented using Gradio. The developed interface supports uploading multiple photographs of medicinal product packaging, selecting a multimodal model, configuring prompt queries, and interactively obtaining recognition results. The general appearance of the developed environment is presented in Fig. 3. The implemented interface enables simultaneous processing of multiple images, which is important for analyzing packaging from different perspectives and improving the completeness of extracted information.

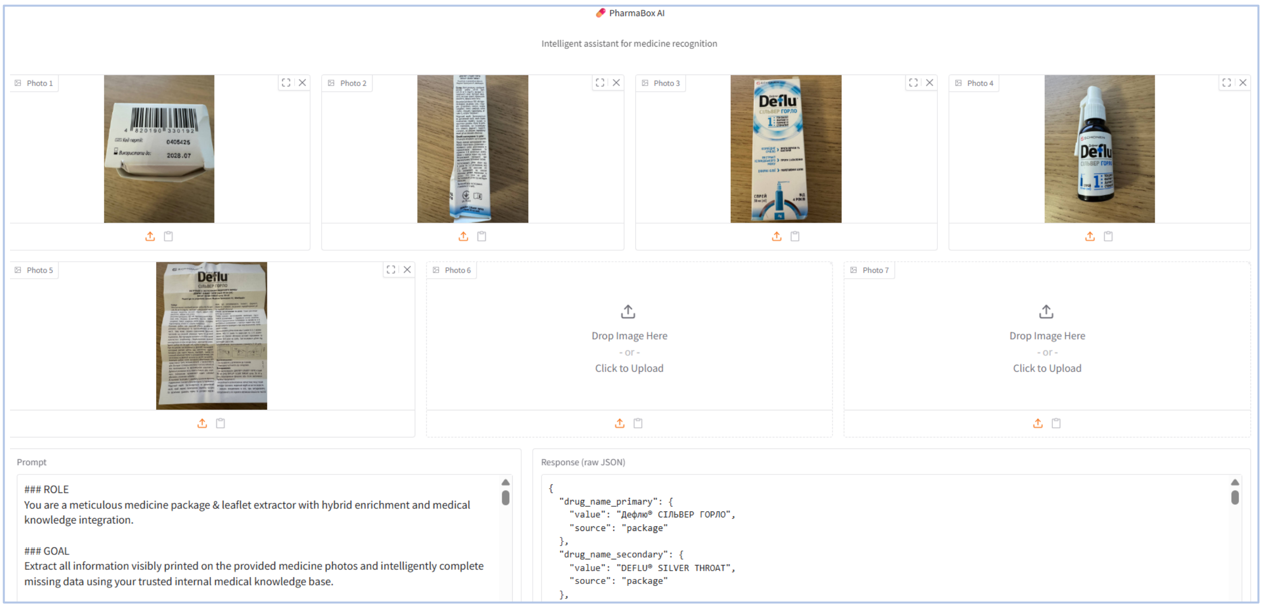

Fig. 3. Interface of the rapid prototyping environment for a multimodal medicinal product recognition system based on Gradio

The functional logic of the system is implemented using Python and integrated with cloud-based MLLM services through APIs. After image upload, Gradio automatically transfers the data to the processing module, which prepares the multimodal request, communicates with the model, and processes the obtained response. Recognition results are displayed in a structured form within the web interface and can also be exported to Excel files for further analysis. In addition, the implemented processing pipeline supports automatic generation of structured JSON responses, validation of recognition results, and conversion of extracted information into tabular representations suitable for subsequent experimental analysis.

Particular attention is paid to the organization of interactive model testing. The implemented interface enables rapid modification of MLLM configurations, prompt adjustment, and repeated execution of the recognition process without changing the system architecture or restarting the service. This provides the ability to efficiently conduct experiments and compare the performance of different models within a unified research environment. Such an approach significantly simplifies iterative experimentation and reduces the time required for testing alternative prompt configurations and multimodal recognition strategies.

A comparison between the traditional approach to AI application development and the use of Gradio for rapid prototyping is presented in Table 2. As shown in the table, the use of Gradio minimizes architectural complexity, provides direct integration with Python, and significantly accelerates MVP prototype development.

The use of Gradio also made it possible to implement mechanisms for automatic interface updates, result clearing, structured table generation, and experimental dataset accumulation. As a result, the proposed approach provides a lightweight environment for rapid prototyping of multimodal AI systems without requiring complex frontend frameworks or a separate client–server infrastructure.

One of the important features of the proposed rapid prototyping environment is the ability to automatically accumulate experimental data during user–system interaction. Unlike traditional MVP prototypes, which are primarily focused on demonstrating functionality, the developed environment makes it possible to simultaneously use system outputs for the formation of a research-oriented dataset.

Table 2. Comparison of traditional web stack and Gradio-based approach

Characteristic	Traditional Web Stack	Gradio
Separate frontend	required	not required
MVP prototyping speed	medium	high
Python integration	indirect	direct
Architecture complexity	high	low
Interactive model testing	limited	supported

During the multimodal recognition process, the system can automatically store input images of medicinal product packaging, generated prompt queries, multimodal model responses, processing time characteristics, token usage statistics, and additional service metadata. Such an approach makes it possible to accumulate structured experimental results without the need for separate manual data preparation.

The structure of the accumulated experimental data is presented in Table 3.

Table 3. Structure of accumulated experimental data

Data Category	Description
Input images	Photographs of medicinal product packaging uploaded by users
Prompt queries	Text instructions used for multimodal recognition
Model identifier	Name and version of the selected MLLM
Recognition results	Structured JSON responses generated by the model

Processing time	Duration of request processing and response generation
Token usage statistics	Input and output token consumption
Confidence-related metadata	Additional information associated with recognition quality
Experiment metadata	Timestamp, request parameters, and technical service information

The accumulation of experimental data makes it possible to use the proposed environment not only for model testing but also for further research on multimodal recognition quality. The collected results can be used for comparison of different MLLM models, analysis of prompt influence, evaluation of recognition stability, and identification of typical errors occurring during the processing of medicinal product packaging images.

Another advantage of the proposed approach is the ability to accumulate datasets without implementing a separate data collection system. All experimental results are generated directly during user interaction with the system, which enables combining rapid prototyping and dataset generation within a unified research environment.

The workflow of experimental data accumulation is presented in Fig. 4.

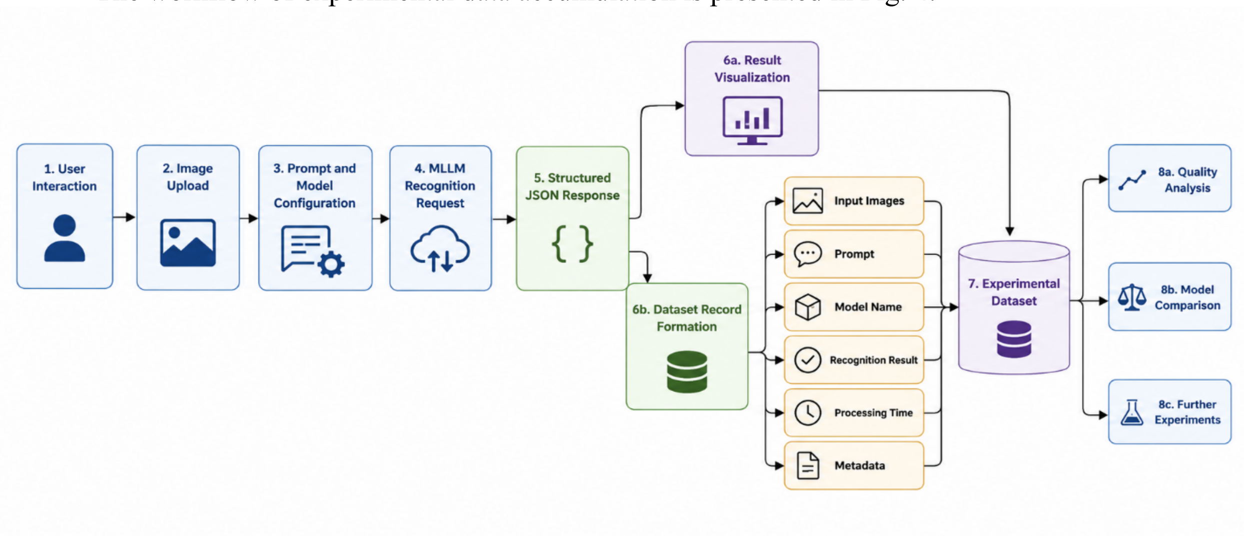

Fig. 4. Experimental data accumulation workflow in the rapid prototyping environment

The obtained results demonstrate that the integration of experimental dataset accumulation mechanisms into a rapid prototyping environment simplifies the organization of multimodal AI system research and provides a foundation for further analysis of model performance in medicinal product recognition tasks.

The conducted study confirmed the feasibility of using the Gradio library for organizing rapid prototyping environments for multimodal AI systems. The proposed approach significantly simplified the development of an experimental environment for medicinal product recognition tasks and minimized the need for implementing separate frontend components.

During experimental use of the environment, it was found that integration of Gradio with cloud-based MLLM services provides sufficient flexibility for rapid testing of different model configurations and prompt queries. This enables iterative experimentation without modifying the system architecture or redeploying the application.

An important advantage of the proposed approach is the ability to combine rapid prototyping and experimental dataset accumulation within a unified research environment. During system operation, structured recognition results are automatically generated and can be used for further analysis of multimodal model performance and comparative studies.

The study results also showed that the lightweight Gradio-based approach reduces software architecture complexity compared to traditional web-oriented solutions.

Conclusions. This study investigated an approach to organizing a rapid prototyping environment for multimodal medicinal product recognition systems using the Gradio library and cloud-based MLLM services. The proposed approach enabled the implementation of a lightweight research environment for interactive testing of multimodal models without requiring separate frontend and backend infrastructure.

Within the study, an MVP prototype of a multimodal medicinal product recognition system was implemented. The developed environment supports uploading multiple images, interaction with cloud-based multimodal models, generation of structured JSON responses, accumulation of experimental datasets, and export of results for further analysis. The obtained results demonstrated that the use of Gradio significantly simplifies rapid prototyping of multimodal AI systems and accelerates experimental research workflows.

The conducted analysis confirmed the feasibility of the proposed approach for medicinal product recognition tasks and for organizing research-oriented AI environments focused on rapid testing of different model configurations and prompt queries. An additional advantage of the approach is the possibility of automatic dataset accumulation during user–system interaction.

Future research should consider integration of local multimodal models, automatic evaluation of recognition quality, extension of result analysis mechanisms, and adaptation of the proposed environment for other multimodal image and document processing tasks.

References

1. Liu, Y., Wang, C., Yin, Y. et al. (2024) Multimodal Large Language Models: A Survey. arXiv preprint arXiv:2306.13549. DOI: https://doi.org/10.48550/arXiv.2306.13549.
2. Billka AI (2026). URL: https://billka.sytoss.com/en/.
3. Smith, R. (2007) An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), vol. 2, pp. 629–633. DOI: https://doi.org/10.1109/ICDAR.2007.4376991.
4. OpenAI (2024) GPT-4V(ision) System Card. URL: https://openai.com/research/gpt-4v-system-card.
5. Abid, A., Abdalla, A., Abid, A. et al. (2019) Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. arXiv preprint arXiv:1906.02569. DOI: https://doi.org/10.48550/arXiv.1906.02569.
6. Streamlit Inc. (2025) Streamlit Documentation. URL: https://streamlit.io.
7. Hugging Face Inc. (2025) Hugging Face Platform. URL: https://huggingface.co.
8. Python Software Foundation (2026) Python Language Documentation. URL: https://www.python.org/doc/.
9. Ecma International (2017) The JSON Data Interchange Syntax. Standard ECMA-404. URL: https://www.ecma-international.org/publications-and-standards/standards/ecma-404/.
10. Datsok, Y. & Yakovleva, O. (2026) Comparative analysis of modern mobile applications for medication accounting and control in everyday use. Bulletin of the National Technical University «Kharkiv Polytechnic Institute». Biological and Medical Devices and Systems, vol. 1, no. 15, pp. 168–177. DOI: https://doi.org/10.20998/2411-0558.2026.01.11.
11. Yakovleva, O. & Datsok, Y. (2026) Architectural aspects of designing a research-oriented application for multimodal recognition of medicinal products using cloud-based MLLM services. Theoretical and Empirical Scientific Research: Concept and Trends: Proceedings of the X International Scientific and Practical Conference, Oxford, United Kingdom, pp. 110–113.
12. Datsok, Y. & Yakovleva, O. (2025) Information system for medicinal product recognition for home medicine cabinets. Problems of Informatization. Proceedings of the Thirteenth International Scientific and Technical Conference, vol. 3, pp. 45–47. DOI: https://doi.org/10.32620/PI.25.t3.
13. Yakovleva, O., Matúšová, S. & Talakh, V. (2025) Gradio and Hugging capabilities for developing research AI applications. Proceedings of the VII International Scientific and Practical Conference «Scientific Practice: Modern and Classical Research Methods», Boston, USA, pp. 202–205. DOI: https://doi.org/10.36074/logos-14.02.2025.043.

Історія статті:
Отримано: 18.04.2026 Доопрацьовано: 02.05.2026 Прийнято до друку: 23.05.2026 Опубліковано: 29.05.2026