DOI: https://doi.org/10.15276/aait.08.2025.7
UDC 004.922

Using large language models for video processing in the agricultural industry

Svitlana G. Antoshchuk[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
Tetiana V. Kunup[1)]
ORCID: https://orcid.org/0000-0003-0246-0951; kunup.t.v @op.edu.ua
Volodymyr I. Lytvynenko[2)]
ORCID: https://orcid.org/0000-0002-1536-5542; immun56@gmail.com. Scopus Author ID: 57189099973
Oleksii V. Danchuk[3)]
ORCID: https://orcid.org/0000-0002-9226-1499; olexdanchuk@gmail.com. Scopus Author ID: 57217231706
[1)] Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
[2)] Kherson National Technical University, Bereslavskoe Shosse 24. Kherson, 73008, Ukraine
[3)] Institute of climate smart agriculture of the national academy of agrarian sciences of Ukraine, 24, Mayatska Doroha, Khlibodarske, Odesa Region, 67667, Ukraine

ABSTRACT

Modern artificial intelligence technologies, particularly large language models, are increasingly being applied in agriculture to enhance automation, decision-making, and sustainability. This study presents a comprehensive analysis of large language models and their integration with computer vision and video processing for real-time livestock monitoring. A software system was developed that utilizes multimodal large language models to analyze poultry behavior from video streams, enabling the detection of anomalies, prediction of potential health issues, and automatic generation of recommendations for farmers. The system is based on a modular architecture and combines technologies such as OpenCV, FastAPI, and Streamlit. Comparative evaluation of models including GPT-4o, Claude 3.7, and LLaVA demonstrates their suitability for different agricultural tasks. The results confirm the effectiveness of large language model-based solutions in improving operational efficiency, reducing human intervention, and supporting precision agriculture. Despite high computational demands, the proposed approach significantly simplifies the deployment of intelligent monitoring systems and opens new opportunities for smart farming innovations.

Keywords: Artificial intelligence; large language models; multimodal models; computer vision; video analysis; poultry monitoring; behavior recognition; precision agriculture; OpenCV; streamlit application

For citation: **Antoshchuk S. G., Kunup T. V., Lytvynenko V. I., Danchuk O. V. "Using large language models for video processing in the agricultural industry"**. ***Applied Aspects of Information Technology*****. 2025; Vol.8 No.1: 88– . DOI: https://doi.org/10.15276/aait.08.2025.7**

1. INTRODUCTION

In the face of global challenges such as climate change, population growth, labor shortages, and the need to enhance food security, the use of advanced information technologies has become a key factor in the development of the agricultural industry. The role of artificial intelligence (AI) and information technology (IT) in agribusiness has been confirmed by the experience of countries such as the United States, Japan, China, and several European nations (Germany, the United Kingdom, the Netherlands, Denmark, and Israel).

Today, new opportunities for industrial development have emerged, driven by the application of large language models (LLMs). Large language models have recently become a sensation in the fields of natural language processing and artificial intelligence and are now developing rapidly. The evolution of LLMs is illustrated in Fig. 1.

Large language models significantly impact various industries by offering intelligent solutions that foster creativity and accelerate the resolution of complex problems. While the most traditional applications of LLMs include automating the creation of articles, reports, and marketing documents, as well as improving translations, their future appears highly promising.

Therefore, research aimed at expanding their capabilities, especially in combination with computer vision technologies and efficiency improvements, is highly relevant. This article explores the potential use of LLMs for processing video data in the agricultural sector.

The aim of this study is to assess the potential and effectiveness of applying large language models in combination with computer vision technologies in agriculture, as well as to demonstrate an example of their efficient use.

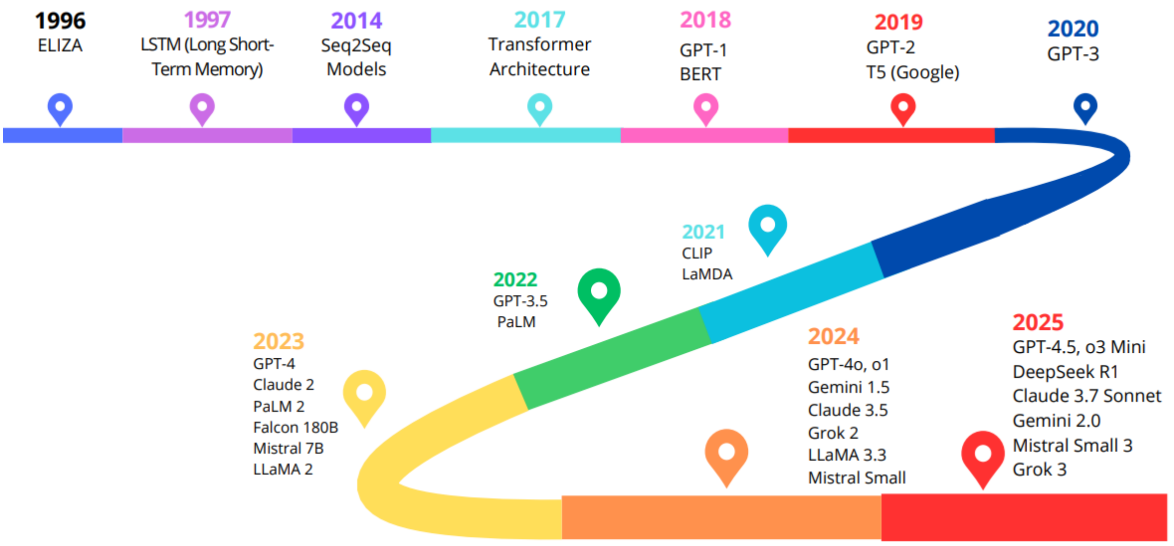

Fig.1 **The evolution of large language models**
Source: **compiled by the authors**

2. ANALYSIS OF LARGE LANGUAGE MODELS

In recent years, natural language processing (NLP) and computer vision (CV) technologies have been rapidly advancing in the field of artificial intelligence, enabling AI systems to learn and reason similarly to humans.

These advancements have led to the emergence of new generations of language models and their multimodal counterparts, such as:

2.1. Statistical Language Models

Early approaches based on methods like n-grams and rule-based systems allowed for modeling the probabilistic distribution of words. While statistical language models **(**SLM) laid the foundation for text processing, their limitations (e.g., lack of contextual understanding) quickly became apparent.

2.2. Neural Language Models (NLM)

With the rise of deep learning, traditional SLMs were replaced by neural network-based models such as feed forward neural networks (FFNN) and recurrent neural networks (RNN). These models could automatically extract features from text, significantly improving language processing quality.

2.3. Pre-trained Language Models (PLM)

The shift toward pre-training enabled the development of universal text representations. There are two main approaches:

Feature-based – The model extracts fixed vector representations, which are then used for specific tasks.

Fine-tuning – A pre-trained model is adapted to the end-user’s tasks, making it more flexible and accurate.

2.4. Large Language Models

Modern large language models **(**LLMs) contain billions of parameters and exhibit so-called “emergent abilities”, allowing them to solve tasks beyond the capabilities of smaller models. Industry leaders include OpenAI's GPT-4.5 and O3, Anthropic's Claude 3.7, Google's Gemini 2.0, DeepSeek's R-1, xAI's Grok-3, Meta's LLaMA 3, and many others.

Due to their scale, LLMs are already being applied across various fields, from data analysis to software development [1, 2], [5, 7], [8, 9], [10, 11], [27, 28], [29, 30].

2.5. Vision-Language Models

Vision-language models (LVM) are systems that integrate the analysis of both textual and visual information. These models can generate detailed image descriptions, answer questions about visual content, and identify hidden semantic relationships. By incorporating data from graphics, audio, and video, LVMs significantly expand the capabilities of traditional NLP systems, enabling comprehensive multimodal data analysis [1, 3].

Thus, the capabilities, number, and diversity of both LLMs and LVMs are continuously growing (see Fig. 1). Large language models have already revolutionized numerous industries, assisting in data analysis, forecasting changes, and optimizing production processes (ComNews). However, it is important to understand that LLMs do not replace experts but rather complement them by accelerating decision-making and improving analysis quality.

Table 1 presents a comparative analysis of large language models that support video information processing [1, 2], [5, 7], [8, 9], [10, 11].

The analysis of the information presented in Table 1 allows us to draw the following conclusions and recommendations regarding the applicability of the models.

GPT-4.5-preview (OpenAI) – Recommended for tasks requiring detailed image analysis with contextual understanding. Its high cost makes it suitable for mission-critical applications where accuracy is paramount [5].

GPT-4 (OpenAI) – Best suited for complex analytical tasks, although it requires significant computational resources [2].

LLaVA2 – Suitable for standalone use and customization but demands high computational power.

Anthropic Claude – Focuses on safety and accuracy in processing scientific data.

PaLM 2 – The best choice for multilingual and scientific tasks, though not as widely available.

For data analysis and reporting, GPT-4, Anthropic Claude, and PaLM 2 are recommended, whereas GPT-3.5 and Llama 2 are better suited for farmer support and request automation. For local use without relying on cloud APIs, LLaVA is a good option. PaLM 2 is ideal for translating research texts and documentation [2, 8], [9].

Table 1. **Comparative characteristics of computer vision models large language models**

Model. (Developer)	Characteristic	Context window. Tokens	Price for 1M tokens
gpt-4.5-preview (OpenAI)	It demonstrates a significant improvement in the interpretation of visual information compared to previous iterations. The model more accurately recognizes subtle details, objects, and context within images, enabling the generation of more precise and detailed descriptions. This is achieved through an optimized visual encoder that transforms images into compact yet information-rich representations [5].	128K	Input: $75.00 Output: $150.00
o1 (OpenAI)	It supports high-detail image input and utilizes additional cross-attention layers for a deeper analysis of visual information. High accuracy is achieved through an additional reasoning step (chain-of-thought), allowing the model to tackle complex visual tasks, although processing may be slightly slower [10].	128K	Input: $15.00 Output: $60.00
GPT-4o (OpenAI)	A flagship model capable of processing high-resolution images; it generates detailed descriptions and answers questions about visual content. It offers high accuracy and stability, with low latency (an average response time of approximately 320ms for voice queries), making it suitable for interactive applications [2].	128K	Input: $2.50 Output: $10.00
GPT-4o-mini (OpenAI)	A lightweight version of GPT-4o with basic image support; it provides adequate recognition but with reduced detail and lower accuracy compared to the full version. It slightly lags behind GPT-4o in image and video processing quality but offers significantly faster processing and lower computational load [1].	32K	Input: $0.15 Output: $0.60
Claude 3.7 Sonnet (Anthropic)	Supports images with enhanced contextual processing. It recognizes complex scenes, objects, and graphics. Accepts images for analysis, including chart and diagram recognition, with improved visual understanding [8].	200K	Input: $3.00 Output: $15.00
Claude 3.5 Haiku (Anthropic)	Basic image support. Well-suited for OCR and document analysis but less effective in interpreting complex scenes. Moderate image recognition accuracy, making it suitable for simplified tasks [9].	200K	Input: $0.80 Output: $4.00
Gemini 2.0 Flash (Google DeepMind)	Accepts multimodal input (text, images, audio, video) and can generate images (native integration with Imagen 3) and audio as output [11].	1M	Input: $0.10 Output: $0.40
DeepSeek R-1 (DeepSeek)	Demonstrates basic image recognition capabilities, making it suitable for tasks that require the integration of textual reasoning with elements of visual analysis while maintaining minimal computational costs [6].	128K	Input: $0.14 Output: $2.19

Source: **compiled by the authors**

DeepSeek R-1 (DeepSeek) – Suitable for tasks with limited computing resources where basic visual analysis needs to be combined with textual reasoning [6].

Thus, vision-language models (LVMs) already provide extensive capabilities for video information analysis on demand, supporting the development and implementation of multimodal models (MLLMs) that integrate text, visual, audio, and video analysis. This will enable even more precise results in complex problem-solving.

It is important to note that the choice of a specific model depends on various factors, including task specificity, budget, infrastructure capabilities, and quality requirements. Next, we will examine the key tasks that can be addressed through the implementation of LLMs in agricultural sector automation.

3. ANALYSIS OF LLM APPLICATIONS IN THE AGRICULTURAL SECTOR

As the literature review shows, large language models (LLMs) can be widely applied in automating routine processes across various industries, including the agricultural sector. Let's consider the key areas where they can be utilized.

Large language models can process vast amounts of data essential for supporting the efficient operation of the agricultural sector, such as analyzing crop yields, assessing soil conditions, and predicting the need for fertilizers and irrigation. They can also be integrated with sensor systems to interpret moisture, temperature, and soil pH level indicators.

Moreover, the agricultural sector has accumulated a vast amount of technical and scientific documentation, regulatory acts, and research reports. Large language models enable the automation of processing this information, facilitating rapid extraction of key data, text analysis, and report generation.

Language models can also analyze economic and meteorological reports, predict changes in supply and demand for agricultural products, and assess the impact of various factors on prices. This helps agricultural companies adapt production and logistics strategies, significantly reducing the workload on specialists and accelerating decision-making processes.

Additionally, the agricultural sector requires multilingual documentation processing, especially in international trade and scientific research. Large language models can automatically translate and adapt complex agronomic texts, simplifying international collaboration and the implementation of advanced technologies.

Farmers, agricultural producers, and distributors can use LLMs for automated customer service, order processing, and the creation of personalized offers based on customer preference analysis [12, 13].

The integration of LLMs with computer vision algorithms (such as OpenCV, YOLO) is a promising direction in modern research practice. This interdisciplinary approach allows for detailed image analysis of crops and livestock, enabling precise identification of symptom [14].

Such analytical approach, washes on processing large volumes of information, facilitates early detection of health issues and enables intervention.

Furthermore, the integration of LLMs into the agricultural sector significantly enhances production efficiency by optimizing management processes, minimizing operational risks, and automating routine tasks. The application of these models supports farmers in making well-informed decisions through comprehensive data analysis and monitoring the health of livestock and crops.

Thus, in the current stage of digitalization, the use of LLMs is gaining strategic importance and becoming a necessary tool for ensuring the sustainable development of agricultural production. A summarized analysis of LLM applications in the agricultural sector is presented in Table 2 [1, 3], [5, 6], [8]. The data in the table highlight the various aspects of LLM use in agriculture, their strengths, and potential limitations.

Large language models can significantly improve the efficiency of various processes in poultry farming, livestock breeding, and other agricultural industries, while also helping to identify key technical and operational requirements for their successful implementation.

Thus, the implementation of large language models and AI in the agricultural sector opens up new opportunities to increase the efficiency and sustainability of the industry, and improve economic indicators.

It should be noted that one of the barriers to the widespread implementation of large language models is the lack of clear determinism in the work of various LLMs, high uncertainty of the final results depending on the data arrays on which the machine learning of the models took place.

Table 2. **Analysis of the Application of large language models in the Agricultural Sector**

Aspect	Application in the agricultural sector	Advantages	Problems/Requirements
Data Analysis and Forecasting	Processing agronomic reports Analysis of sensor data and market trends [16]	Improved decision making Improved yield predictions Optimized resource use	Large, high-quality data sets are required The need to create reliable data feeds
Automated monitoring	Animal Behavior Analysis via Video Streams Health Indicator Monitoring [17]	Early detection of illness and signs of stress Reducing dependence on manual monitoring	Integration with computer vision systems High demands on real-time processing
Decision support	Providing prompt recommendations through virtual assistants and chatbots [18, 19], [20, 21]	Timely recommendations for feeding, treatment and management Improving operational efficiency	Requires customization for specific agricultural conditions Expert interpretation of results required
Document processing	Automation of translation, summarization and analysis of technical documentation [22, 23]	Optimize compliance with regulations and standards Multilingual support Quick access to key information	Complexity of technical language Ensuring data security and accuracy
Customer interaction and support	Virtual assistants for farmer consultations and processing requests [25]	Improved support for farmers Quick resolution of problems Increased availability of expert advice	Need for training in a specific agricultural context Ensuring reliability and accuracy of recommendations
Integration with IoT and sensors	Combining LLM analytics with IoT device data (e.g. environmental sensors, wearables) [15, 24], [26]	Holistic view of farm operations Optimized environmental and resource management	Significant investment in infrastructure Ensuring compatibility and interoperability

Source: **compiled by the authors**

These circumstances are also aggravated by the general availability of most LLMs and the insufficient qualifications of most categories of users.

Let us consider the implementation of multimodal models (MLLM) to solve the problem of monitoring chickens.

4. LARGE LANGUAGE MODELS IN VIDEO ANALYSIS

Today, automation of data collection and analysis processes is an important storage technology for human intelligence and computer vision. We will evaluate the potential and effectiveness of multimodal LLMs for bird welfare monitoring.

As is well known, the use of large language models (LLMs) in text processing has already been well established, and there are several recommendations for prompt engineering, where instructions or queries are provided to the model to fine-tune its interaction with generative AI. However, video analysis requires a more complex approach, including additional preparation and management stages, for which an information system has been developed, as shown in Fig. 2.

These stages can be broadly divided into four main groups.

1. Split Video into Frames – Video Processing and Segmentation into Fragments

Before an LLM can analyze a video, it must be converted into a format suitable for processing. This process involves several steps.

Splitting the video into frames – The video is represented as a sequence of images. In our system, the segmentation is performed based on a time interval specified by the user.

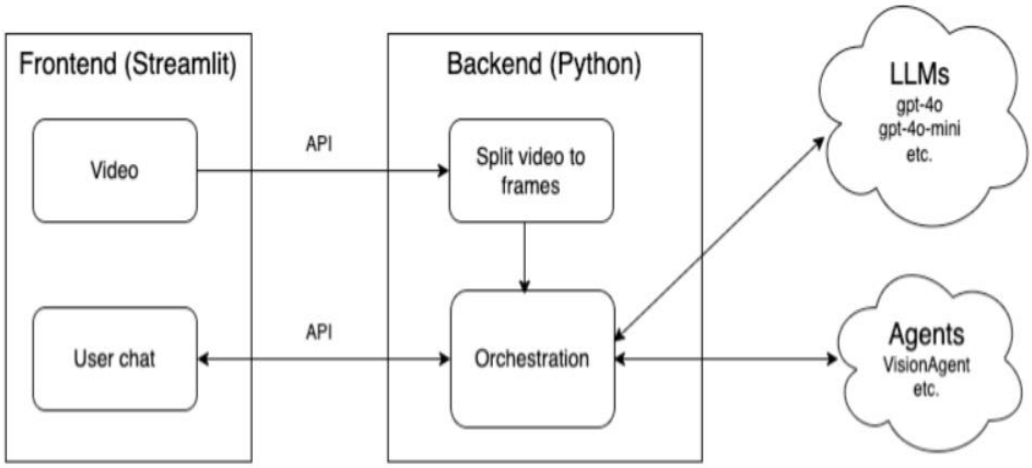

Fig. 2. **Program structure**
Source: **compiled by the authors**

Adding metadata to images – Each frame is assigned metadata, including a timestamp and other technical information. This is necessary for further analysis, synchronization with other data, and the correct interpretation of results.

Preprocessing images – This includes enhancing image quality: improving sharpness, balancing brightness and contrast, and reducing noise, all of which help increase the accuracy of the analysis.

2. User Chat – User Interaction Interface

The user interacts with the system through the User Chat, entering text-based queries and specifying processing parameters, language settings, the choice of LLM, and other options. This interface not only allows users to define tasks, such as "Identify what is in the frame and describe it", but also enables continued interaction in a chat-based format.

3. Video Processing – Video analysis and processing

This stage is responsible for detailed analysis of video content taking into account the user's input request, as well as for coordinating the interaction of various system modules.

After splitting the video into frames, the system performs:

Information extraction – object detection, face recognition, text extraction, analysis of emotions and other visual characteristics.

Scene description generation – LLM builds meaningful text descriptions of what is happening in the video. For example: "In the video, chickens are in a poultry house. Most of them are white".

Search and classification of objects - identification of various elements, such as chickens, drinkers, feeders and other objects.

Event extraction – analysis of time series to determine key moments, including detection of dangerous situations or recording of behavioral anomalies.

4. Orchestration – orchestration of interactions between system components

For efficient operation, the system coordinates the work of the modules:

Synchronization of CV and NLP – computer vision (CV) algorithms recognize objects in the frame, and LLM generates a text description.

Query optimization (prompts) – if the user's request is complex, it is automatically divided into subtasks for step-by-step processing.

Feedback from the user – the system generates intermediate reports and clarifies the details of the task to improve the accuracy of processing.

The following technologies and libraries were used in the construction of the IS.

Web Framework and Server

FastAPI – a high-performance web framework for building APIs in Python with asynchronous support.

Uvicorn – an ASGI server used to run FastAPI applications.

File and Multimedia Processing

Python-multipart – a library for handling multipart/form-data, enabling file uploads via API.

YT-DLP – a tool for downloading videos from YouTube and other platforms.

OpenCV – a library for image and video processing.

Pillow – an extended version of the PIL library, used for image manipulation.

Data Validation and Security

Pydantic – a library for data validation and schema management.

Python-JOSE – a library for working with JSON Web Tokens (JWT).

Passlib [bcrypt] – a library for secure password hashing.

Database and Migrations

SQLAlchemy – a powerful Object-Relational Mapping (ORM) tool for database management.

Alembic – a tool for managing database schema migrations.

UI Development and API Interaction

Streamlit – a library for creating interactive web applications with minimal code.

HTTPX – An asynchronous client for HTTP requests, used for interacting with external APIs.

Python-dotenv – a tool for loading environment variables from .env files.

Pydantic-settings – a Pydantic extension for managing application settings.

Testing

Pytest – a testing framework that supports parameterization and unit testing [31].

Artificial Intelligence

OpenAI (1.6) – a client for interacting with OpenAI API (ChatGPT-4, GPT-4-mini).

These technologies and libraries were ideal for developing a system that includes video processing, API integration, user management, and testing. They comprehensively cover all key requirements of the information system (IS) – from database and security to external service interaction and frontend development.

Fig. 2 illustrates the general structure of the system, which consists of three main elements: Frontend (Streamlit), Backend (Python) and external services in the form of LLMs (large language models) and Agents (specialized modules).

Frontend (Streamlit): interacts with the backend via API: sends requests for video processing and receives results or answers to questions.

Backend (Python): consists of two main components:

Split video into frames: a module that splits the video into frames at given intervals.

Orchestration: a central logical unit that coordinates the entire system. It receives requests from the Frontend, calls the appropriate services (video processing modules, LLMs, agents, etc.) and compiles the response to send back to the Frontend. It makes calls to external services (LLMs and agents) to solve various tasks: image analysis, text response generation and anomaly detection.

LLM (Large Language Models): These models can perform tasks such as text generation, description and image analysis, and answering user questions. They are called by the backend to process text queries and images extracted from videos.

Agents (VisionAgent): Specialized modules or microservices that perform specific functions related to image or video processing and analysis. They can use computer vision, machine learning, or other algorithms to detect anomalies, classify objects, and evaluate chicken behavior. Its main functions include:

Anomaly Detection – VisionAgent analyzes frames extracted from videos to detect abnormal behavior patterns, such as signs of distress or unusual movements of chickens Object Classification – It uses machine learning algorithms to classify objects in video frames, identify key characteristics, and distinguish between different elements in a scene.

Behavior Analysis – Using computer vision techniques, VisionAgent helps evaluate chicken behavior, supporting the identification of signals of stress, aggression, or deviations from natural activity patterns.

Integration with other services – VisionAgent is designed to seamlessly integrate with the backend of the system and other artificial intelligence components, allowing it to work in tandem with language models and orchestration modules to provide comprehensive information.

Real-time processing – Optimized for performance, VisionAgent can process video frames in real-time, facilitating timely monitoring and decision-making in automated systems.

These features make VisionAgent a key component for implementing robust automated video analysis solutions in applications such as chicken behavior monitoring and animal welfare in poultry farming.

The diagram illustrates how the interface (Frontend), data processing logic (Backend), and external intelligent services (LLMs, Agents) interact to effectively analyze chicken video and behavior.

The interaction between modules is shown in Fig. 2 and was previously described.

The system allows users to upload videos from video services or local files, choose a communication language, and utilize large language models. Additionally, it provides the option to configure video segmentation parameters by specifying the interval at which frames (images extracted from the video) are generated – that is, setting the number of seconds between frames.

Fig. 3 shows the main page of the system. Left Panel: "Upload Options" (Fig. 3).

Choose upload method: The user can select the source for uploading the video (either a local file or a YouTube link).

Select frame interval (seconds): Specify the interval in seconds at whom frames will be generated from the video.

Choose language: Allows the user to choose the language for communication or analysis.

Select model: Enables selection of a large language model (LLM) to be used for further processing.

Right Panel: "Video Processing Dashboard" (Fig. 3).

Drag and drop file here / Browse files: An area where the file can be dragged and dropped for uploading or selected from local storage.

Video Title: A field displaying the title of the uploaded video (e.g., "farm-chickens-1.mov") and its file size.

Description: A text field for additional information or a description of the video (e.g., "Poultry farm").

Process Video: A button that initiates video processing (splitting the video into frames, passing the frames to the selected LLM).

This interface is built on the Streamlit framework, which allows for the rapid and convenient development of web applications for processing and analysis video. The backend is implemented using the Python programming language, and communication between the interface and the server is carried out via a RESTful API.

The system development uses several powerful technologies Fig.4.

FastAPI is a modern web platform for creating high-performance RESTful APIs. Thanks to its asynchronous capabilities, FastAPI provides fast request processing and easy scaling.

OpenCV is a library for image and video processing. It allows you to extract frames from videos, resize them, and apply filters and algorithms for analysis.

Streamlit is a library for rapid development of interactive web applications for scientific research and data analysis. It provides a convenient interface for interacting with the processing results.

The system has a modular architecture that includes the following components: API (FastAPI) – The API (FastAPI) is responsible for processing requests and interacting with the user via the REST API. This allows you to load videos, select the frame extraction interval, and access the processing results.

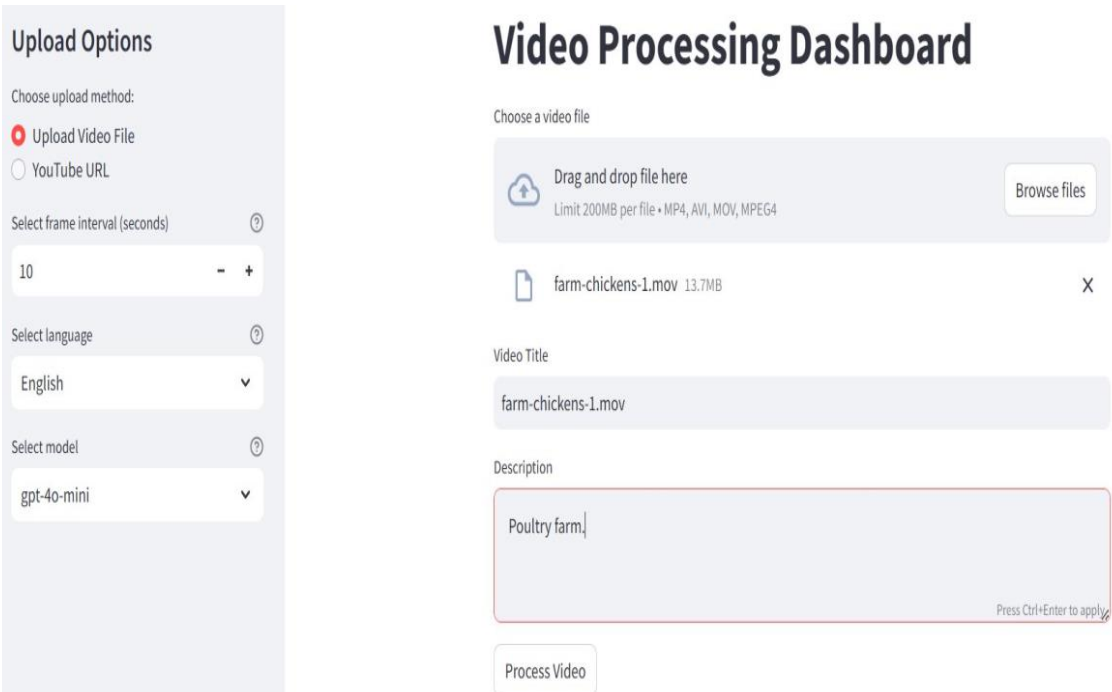

Fig. 3. **Screenshot of the main page of the program**
Source: **compiled by the authors**

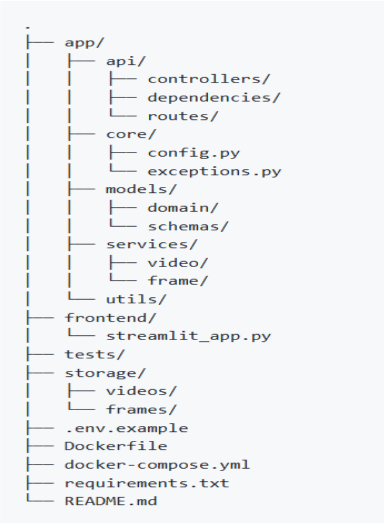

Fig. 4. **Screenshot of the system structure**
Source: **compiled by the authors**

The video processing service (OpenCV) is a module that performs basic video operations, such as extracting frames at specified intervals and saving the resulting images.

The interface (Streamlit) creates a user interface where users can upload videos, select intervals for analysis, and view the results in real time. Figure 4 shows a screenshot of the system structure.

5. VIDEO PROCESSING AND FRAME ANALYSIS METHODS

A video processing system is presented that splits the video stream into individual frames and analyzes them. In this example, the user has selected processing parameters, including the upload method (a local file, but a YouTube link can also be chosen), the frame extraction interval set to 10 seconds, the analysis language (English), and the processing model (gpt-4o-mini).

Video Upload. The user can upload videos from a local computer or insert a YouTube link. Video processing in various formats is supported, which provides flexibility in working with data.

Frame Extraction. The video processing service selects frames from a video at a specified time interval, the interval is set on demand (from 1 to 20 seconds). This is implemented by calculating the required number of frames in accordance with the video duration and the specified interval. At the same time, it is possible to adapt to specific analysis scenarios, for example, to detect dynamic changes or monitor the activity of objects.

Fig. 5 shows the processing status, which confirms the successful completion of the process and the extraction of three frames from the video.

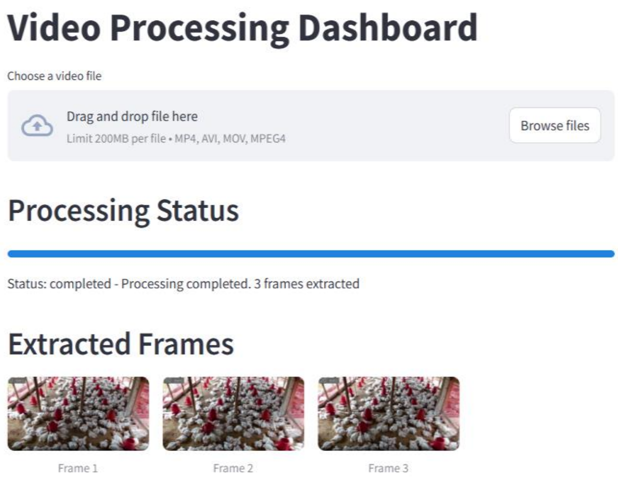

Fig. 5. **Screenshot of selecting frames from a video at a specific time interval**
Source: **compiled by the authors**

Frame Analysis. The resulting footage can be further processed, such as filtering, colour analysis, and machine learning algorithms can be used to detect anomalies in the behaviour of chickens. Additionally, object recognition, motion tracking and detection of changes in behaviour based on historical data are supported. This allows for the effective identification of potential risks, such as stress or aggressive behaviour in birds.

Interactive Interface. Through the interface, users can view processed frames, ask questions, clarify details, and analyze individual frames from the video. Additionally, there is an option to configure processing parameters, such as selective highlighting of frame areas, adjusting contrast and detail levels, and generating automated reports based on video analysis.

The system output is presented as three frames extracted at a specified interval. The automatically generated description indicates that the video depicts a poultry farm with a large number of white chickens freely moving across a floor covered with sawdust or straw, creating a natural environment for their movement (Fig. 6).

The system interface also allows users to manually describe frames, providing a corresponding input field. The automatically generated description, displayed in a yellow box, is based on scene analysis using the LLM model GPT-4o-mini.

Overall, the system successfully processed the video, extracted key frames, and generated textual descriptions, simplifying video data analysis and providing a convenient tool for working with visual information.

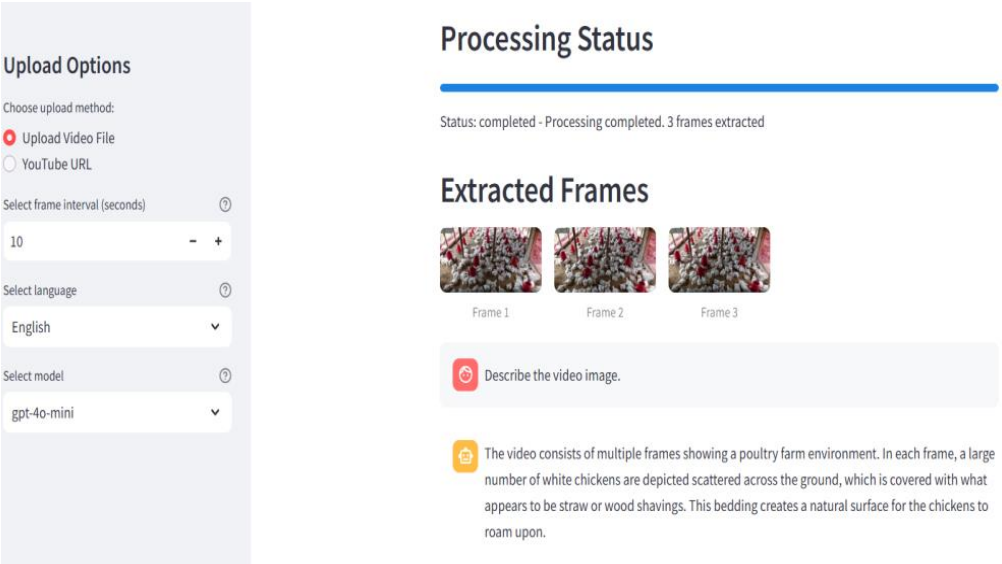

Fig. 6. **Screenshot of selecting frames from a video at a specific time interval**

Source: **compiled by the authors**

The developed system utilizes multimodal LLMs from OpenAI to analyze chicken behavior based on video streams, enabling the detection of anomalies in their activity, predicting potential health issues, and automatically generating recommendations for farmers.

The study confirms that the integration of LLMs in agriculture ensures automated animal monitoring, reduces reliance on human factors, and supports real-time decision-making.

CONCLUSION

Studies have confirmed that the integration of large language models (LLMs) into industry, including the agricultural sector, contributes to increased production efficiency, automation of routine processes, and improved decision-making quality.

Traditional computer vision models, such as YOLO (You Only Look Once), require significant resources for development and maintenance. The process involves collecting extensive image datasets, meticulously annotating them, and continuously updating models to adapt to changing conditions. This necessitates the involvement of machine learning specialists and annotators, increasing costs and time expenditures.

With the advent of large language models (LLMs) and vision-language models (LVMs), the development and implementation of artificial intelligence systems have become more accessible and flexible. These models are trained on vast amounts of data, allowing them to process both textual and visual information without requiring specialized datasets or complex annotation. Large language models and LVMs can perform a wide range of tasks, from text generation to image analysis, making them versatile tools for various industries.

Although the use of LLMs and LVMs is currently expensive due to high computational requirements and licensing costs, it is expected that their cost will decrease over time. Advances in technology and the increasing number of open-source models are expanding the accessibility of these tools to a broader audience. This will enable companies and organizations to leverage cutting-edge AI technologies without substantial financial investments, reducing barriers to adoption and innovation. A comparative analysis of different LLMs, including GPT-4, Flamingo, and LLaVA, highlights their potential in video processing and multimodal data analysis.

Thus, the transition from specialized models like YOLO to more universal LLMs and LVMs provides significant advantages in flexibility, scalability, and cost reduction for the development and maintenance of AI systems.

The use of large language models (LLM) in video analysis opens up new possibilities for automation and increased accuracy of video data analysis. However, like any technology, this system has both advantages and disadvantages.

Advantages. Automation and reduced human workload. The system can automatically analyze video without requiring human intervention at each stage. It allows processing large volumes of video data faster than manual analysis. Reduces costs associated with human resources, especially in the field of security, monitoring and analytics. LLMs do not just recognize objects, but also generate meaningful text descriptions of events. They can identify complex behavior patterns, such as incident analysis or detection of potentially dangerous situations. The system can be adapted for various tasks, including video surveillance, quality control and health monitoring. Real-time processing. Thanks to optimized architectures, video can be processed in real time, which is useful for surveillance and production monitoring. Increased accuracy of analysis. The combination of LLM and computer vision improves the accuracy of video analysis. LLMs can help with contextual interpretation, which traditional computer vision algorithms cannot do. Integration with various data sources. The ability to combine data from different sensors (cameras, temperature sensors, humidity sensors, and other IoT devices). Support for multimodal data, including text, images, audio, and video, for deeper and more accurate analytics.

Disadvantages of video processing systems with large language models. The main drawback of a video processing system using LLMs is the high computational load, as it requires powerful hardware, including servers and GPUs or TPUs, especially when performing real-time analysis. The combination of LLMs with computer vision significantly increases resource consumption, leading to high data processing costs.

Another issue is the long processing time for large volumes of data. Splitting video into frames, analyzing them, and generating textual descriptions require substantial computational power, particularly for long video recordings. Additionally, some models show low efficiency when working with low-quality video, which can reduce analysis accuracy.

The high cost of hardware, including powerful servers and storage solutions, as well as expenses for API access to models, is another challenge. However, this disadvantage is gradually becoming less significant as the cost of computational resources and API access decreases, making the system more accessible.

Based on these considerations, it can be concluded that the system demonstrates high efficiency in video analysis, enabling automation of data processing and interpretation. However, further improvements are needed, particularly in optimizing computational processes, enhancing result interpretation, and ensuring data privacy. Addressing these aspects will improve performance, reduce hardware load, and enhance information security, making the system more robust and adaptable to various operational conditions.

REFERENCES

1. “GPT-4o mini: advancing cost-efficient intelligence”. – Available from: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence. – [Accessed: Nov. 2024].

2. Cheung, D. “Meta Llama 2 vs. OpenAI GPT-4: A comparative analysis of an open-source vs. proprietary LLM”. – Available from: https://www.codesmith.io/blog/meta-llama-2-vs-openai-gpt-4-a-comparative-analysis-of-an-open-source-vs-proprietary-llm. – [Accessed: Nov. 2024].

3. “Independent analysis of AI models and API providers”. – Available from: https://artificialanalysis.ai. – [Accessed: Nov. 2024].

4. “Seq2seq”. – Available from: https://github.com/farizrahman4u/seq2seq/blob/master/README.md. – [Accessed: Nov. 2024].

5. “Introducing GPT-4.5”. – Available from: https://openai.com/index/introducing-gpt-4-5/ – [Accessed: Nov. 2024].

6. “DeepSeek-V3 Capabilities”. – Available from: https://www.deepseek.com. – [Accessed: Nov. 2024].

7. “Hello GPT-4o”. – Available from: https://openai.com/index/hello-gpt-4o. – [Accessed: Nov. 2024].

8. “AI research and products that put safety at the frontier”. – Available from: https://www.anthropic.com. – [Accessed: Nov 2024].

9. “Claude 3.5 Haiku”. – Available from: https://www.anthropic.com/claude/haiku. – [Accessed: Nov. 2024].

10. “Introducing OpenAI o1”. – Available from: https://openai.com/o1. – [Accessed: Nov. 2024].

11. “Gemini 2.0 Flash”. – Available from: https://deepmind.google/technologies/gemini/flash. – [Accessed: Nov 2024].

12. Wu, J., Lai, Z., Chen, S., Tao, R., Zhao, P. & Hovakimyan, N. “The new agronomists: Language models are experts in crop management”. *arXiv*. 2024. DOI: https://doi.org/10.48550/arXiv.2403.19839.

13. Bist, R. B., Subedi, S., Chai, L. & Yang, X. “Ammonia emissions, impacts, and mitigation strategies for poultry production: A critical review”. *Journal of Environmental Management*. 2022; 8 (2): 1–14. DOI: https://doi.org/10.1016/j.jenvman.2022.116919.

14. Silva, B., Nunes, L., Estevão, R., Aski, V. & Chandra, R. “GPT-4 as an agronomist assistant? Answering agriculture exams using large language models”. 2023. DOI: https://doi.org/10.48550/arXiv.2310.06225.

15. Chai, L., Xin, H., Wang, Y., Oliveira, J., Wang, K. & Zhao, Y. “Mitigating particulate matter emissions of a commercial cage-free aviary hen house”. *ASABE Annual International Meeting*. Detroit, USA. 2018. p. 1800223. DOI: https://doi.org/10.13031/aim.201800223.

16. Castro, F., Chai, L., Arango, J., Owens, C., Smith, P., Reichelt, S., DuBois, C. & Menconi, A. “Poultry industry paradigms: connecting the dots”. *Journal of Applied Poultry Research*. 2023; 32 (1): 100310. DOI: https://doi.org/10.1016/j.japr.2022.100310.

17. Fernandez, A. P., Norton, T., Tullo, E., van Hertem, T., Youssef, A., Exadaktylos, V., Vranken, E., Guarino, M. & Berckmans, D. “Real-time monitoring of broiler flock’s welfare status using camerabased technology”. *Biosystems Engineering*. 2018; 173 (2): 103–114. DOI: https://doi.org/10.1016/ j.biosystemseng. 2018.05.008.

18. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J. & Poon, H. “Domain-specific language model pretraining for biomedical natural language processing”. *ACM Transactions on Computing for Healthcare (HEALTH)*. 2021; 3 (1): 1–23. DOI: https://doi.org/10.1145/3458754.

19. Zhuang, F., Xiong, H., Zhang, Q., Guo, Q. et al. “A survey on knowledge graph-based recommender systems”. *IEEE Transactions on Knowledge and Data Engineering*. 2020; 34; 937. DOI: https://doi.org/10.1360/SSI-2019-0274.

20. Liang, C., Berant, J., Le, Q., Forbus, K. & Lao, N. “Neural symbolic machines: Learning semantic parsers on freebase with weak supervision”. *1st Workshop on Neural Abstract Machines & Program Induction (NAMPI)*. Barcelona, Spain. 2016. p. 23–33. DOI: https://doi.org/10.48550/arXiv.1612.01197.

21. Lu, G., Li, S., Mai, G. et al. “AGI for Agriculture”. *arXiv*. 2023. p. 1–18. DOI: https://doi.org/10.48550/arXiv.2304.06136.

22. Kuchmijova, T. S. “Digitalization of business processes in the conditions of transformational transformations changes” [In Ukrainian]. *Investytsii: Praktyka and Dosvid*. 2023; 10: 77–80. DOI: https://doi.org/10.32702/2306-6814.2023.10.77.

23. “FarmVibes.AI: Multi-Modal GeoSpatial ML Models for Agriculture and Sustainability”. 2021. – Available from: https://github.com/microsoft/farmvibes-ai. – [Accessed: Nov. 2024].

24. Patel, H. & Sana, A. “Role of computer science (Artificial Intelligence) in poultry management”. *Devotion Journal of Research and Community Service*. 2022; 3 (12): 2068–2088. DOI: https://doi.org/10.36418/dev.v3i12.250.

25. Poliakov, O. “Artificial intelligence in agricultural production” [in Ukrainian]. *Materials of the International Scientific and Practical Conference Youth and Technical Progress in A*PV. *Innovative Developments in the Agricultural Sector*. Kharkiv, Ukraine. 2021; 2: 408–409. – Available from: https://khntusg.com.ua/wp-content/uploads/2020/02/molod_2021.pdf#page=410. – [Accessed: Nov. 2024].

26. Schlegel, L., Kleine, A. S., Doherr, M. G. & Fischer-Tenhagen, C. “How to see stress in chickens: On the way to a Stressed Chicken Scale”. *Poultry Science*. 2024; 103 (8): 103875. DOI: https://doi.org/10.1016/j.psj.2024.103875.

27. Shevchenko, A. I., Baranovsky, S. V., Bilokobylsky, O. V., et al. “Strategy for the development of artificial intelligence in Ukraine”. Monograph / [Edited by A.I. Shevchenko]. Kyiv: IISS, 2023. 305 c.
ISBN: 978-617-7894-89-5. DOI: https://doi.org/10.15407/development_strategy_2023.

28. Wang. X., Wei. J. & Schuurmans. D. et al. "Self-consistency improves chain of thought reasoning in language models". *The Eleventh International Conference on Learning Representations*. 2023. DOI: https://doi.org/10.48550/arXiv.2203.11171.

29. Wei, J., Wang, X., Schuurmans, D. et al. "Chain-of-thought prompting elicits reasoning in large language models". *36th Conference on Neural Information Processing Systems*. 2022. DOI: https://doi.org/10.48550/arXiv.2201.11903.

30. Wei, J, Tay, Y, Bommasani, R. et al. "Emergent abilities of large language models". *ArXiv*. 2022. DOI: https://doi.org/10.48550/arXiv.2206.07682.

31. Antoshchuk, S. G. & Breskina, A. A. "Evaluation metrics systematization for 2D human poses analysis models". *Applied Aspects of Information Technology. Publ. Nauka i Tekhnika*. Odessa: Ukraine. 2023; Vol.6, No.1: 26–38. DOI: https://doi.org/10.15276/hait.06.2023.2.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article
Author SvitlanaG.Antoshchuk is a member of the Editorial Board ofthis journal.This role had no influence on the peer review process or editorial decision regarding this manuscript

Received 27.12.2024
Received after revision 14.03.2025
Accepted 19.03.2025

DOI: https://doi.org/10.15276/aait.08.2025.7
УДК 004.922

Використання великих мовних моделей в аграрній промисловості

Антощук Світлана Григорівна[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
Кунуп Тетяна Василівна[1)]
ORCID: https://orcid.org/0000-0003-0246-0951; kunup.t.v@op.edu.ua
Литвиненко Володимир Іванович[2)]
ORCID: https://orcid.org/0000-0002-1536-5542; immun56@gmail.com. Scopus Author ID: 57189099973
Данчук Олексій Володимирович[3)]
ORCID: https://orcid.org/0000-0002-9226-1499; olexdanchuk@gmail.com. Scopus Author ID: 57217231706
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна
[2)] Херсонський Національний технічний університет, Береславське шоссе 24. Херсон, 73008, Україна
[3)] Інститут кліматично орієнтованого сільського господарства НААН України, вул. Маяцька дорога, 24, смт Хлібодарське, Одеська область, 67667, Україна

АНОТАЦІЯ

Сучасні технології штучного інтелекту (ІІ) та великі мовні моделі (LLMs, Large Language Models) активно впроваджуються в різні галузі промисловості, у тому числі і в аграрний сектор, сприяючи автоматизації моніторингу, підвищенню продуктивності та сталого розвитку. У цій статті проведено аналіз LLMs. Показано їх переваги для аналізу даних, прогнозування змін та оптимізації виробничих процесів, особливо при їх інтеграції з комп'ютерним зором та технологіями обробки відеоданих. Відзначено перспективи застосування в агросекторі (у птахівництві, тваринництві та ін.) Розроблена система використовує мультимодальні LLMs для аналізу поведінки курей на основі відеопотоків, що дозволяє виявляти аномалії в їхній активності, передбачати можливі проблеми зі здоров'ям та автоматично генерувати рекомендації для фермерів. Дослідження підтверджує, що впровадження LLMs у сільське господарство забезпечує: автоматизований моніторинг тварин та сільськогосподарських культур, покращення точності прогнозів урожайності та стану ґрунту, зниження залежності від людського фактору, підтримку прийняття рішень у режимі реального часу. Порівняльний аналіз різних моделей LLMs, GPT-4, Flamingo та LLaVA демонструє їх потенціал у обробці відео та мультимодальному аналізі даних. Представлені результати підтверджують, що використання LLMs у поєднанні з технологіями машинного навчання та комп'ютерного зору відкриває нові перспективи для точного землеробства та автоматизованого контролю за здоров'ям тварин, роблячи сільськогосподарське виробництво більш технологічним, економічним та екологічним.

Ключові слова: штучний інтелект; великі мовні моделі; комп'ютерний зір; моніторинг, LLM; птахівництво; OpenCV; мультимодальні моделі

ABOUT THE AUTHORS

Svitlana G. Antoshchuk - Doctor of Engineering Sciences, Professor, Department Information Systems.
Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
***Research field*:** Automated management systems and progressive information technologies; information technology; theoretical and applied aspects of multimedia data processing

Антощук Світлана Григорівна - доктор технічних наук, професор кафедри Інформаційних систем.
Національний університет «Одеська політехніка», пр. Шевченка, 1, Одеса, 65044, Україна

Tetiana V. Kunup - PhD, Department of Information Technologies. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/ 0000-0003-0246-0951; kunup.t.v @op.edu.ua
***Research field*:** Information technology

Кунуп Тетяна Василівна - кандидат технічних наук, ст. викладач кафедри Информационніх технологий.
Національний університет «Одеська політехніка», пр. Шевченка, 1, Одеса, 65044, Україна

Volodymyr I. Lytvynenko - Doctor of Engineering Sciences, Professor, Head of Department Informatic and Computer Science. Kherson National Technical University, 24, Bereslavskoe Shosse. Kherson, 73008, Ukraine
ORCID: https://orcid.org/0000-0002-1536-5542; immun56@gmail.com. Scopus Author ID: 57189099973
***Research field*:** Artificial intelligence; large language models; multimodal models; machine learning; computer vision

Литвиненко Володимир Іванович - доктор технічних наук, професор завідувач кафедри Інформатики і компьютерних наук Херсонського Національного технічного університету, Береславське шоссе 24. Херсон, 73008, Україна

Oleksii V. Danchuk - Doctor of Veterinary Sciences, Professor, Deputy Director for Scientific Work.
Institute of climate smart agriculture of the National academy of agrarian sciences of Ukraine,
Mayatska Doroha 24, Khlibodarske. Odesa Region, 67667, Ukraine
ORCID: https://orcid.org/0000-0002-9226-1499; olexdanchuk@gmail.com. Scopus Author ID: 57217231706
***Research field*:** Neurohumoral regulation of physiological functions, nanotechnology, implementation of digital technologies in animal husbandry and veterinary medicine

Данчук Олексій Володимирович - доктор ветеринарних наук, професор, заступник директора з наукової роботи Інституту кліматично орієнтованого сільського господарства НААН України, вул. Маяцька дорога, 24, смт Хлібодарське. Одеська область, 67667, Україна