

How to Successfully Integrate Computer Vision, Large Language Models, and Intelligent Document Processing

A use case with ABBYY Vantage



How to successfully integrate computer vision, large language models, and intelligent document processing

Hyperautomation is an approach used in organizations to quickly identify, verify, and automate as many processes as possible. In this context, intelligent document processing (IDP) is certainly a key component, since most organizational processes are based on electronic documents or paper.

IDP is capable of capturing, extracting, and processing data from a variety of document formats.

An IDP solution applies artificial intelligence (AI) and machine learning (ML) techniques to process structured, semi-structured, and unstructured documents. These technologies understand the content of documents and are able to extract the data, allowing for process automation.

However, with the advancement of the use of new technologies, some business processes move from only documents to images, videos, and document interpretation. In this scenario, additional technologies are needed to achieve hyperautomation.

In business processes that include images and videos, the use of computer vision (CV) is more efficient than IDP technology. **Computer vision is an area of artificial intelligence that focuses on the ability of machines to interpret and understand visual content and can be trained to recognize a wide variety of objects, patterns, and features in images.** The use of CV can be added to the automation process by consuming services such as Microsoft Cognitive Services, AWS Amazon Rekognition, and Google Cloud Vision, among others, or even through specialized open libraries such as OpenCV and Python programming language.

Similarly, there are scenarios that go beyond data extraction, in which an LLM can contribute to a higher level of automation. **Large language models (LLMs) are an area of artificial intelligence that use algorithms to create human language from a large set of texts used as reference.** If the task involves understanding of complex contexts, activities such as document summarization, translation, or simply answering a question related to the context, an LLM may be more suitable. OpenAI (ChatGPT), Microsoft (Bing), and Google (Bart) are services that can be integrated into IDP solutions to perform these activities.



Using **ABBYY Vantage**, our low-code intelligent document processing platform, it is possible to integrate technologies such as CV and LLM into the platform to obtain an automated process with the least level of human interaction possible. Integrating these technologies allows for a multidisciplinary approach to automation, dealing with different types of data and information.

Claims automation: An ABBYY Vantage use case

The ABBYY Vantage intelligent document processing platform provides easily consumable artificial intelligence-based skills/models to understand documents quickly and easily.

Vantage provides over 150 pre-trained skills (AI models) that can read, understand, and extract information from business documents, helping businesses accelerate hyperautomation.

Additional technologies such as CV and LLM can be easily consumed by ABBYY Vantage with API execution. Serverless components like Azure Functions, AWS Lambda, or Google API Engine are flexible, scalable, and pay-as-you-go—often a great method for incorporating technologies.

In this practical insurance use case, we created a process for claims automation, in which various documents can be sent, such as vehicle photos, odometer photos, driver's licenses, police reports, and invoices through the ABBYY Vantage mobile capture application.



The photos of the vehicles must be validated according to the position of the vehicle, and must be taken front left, front right, rear left, and rear right with the purpose of proving the condition of the vehicle. The odometer photo, on the other hand, is intended to prove the current mileage of the vehicle. Driver's license, police report, and invoice aim to provide complementary information on the process.

VIDEO

Claims Automation with

Intelligent Document Processing

An ABBYY Vantage Use Case

Description of activities:



Mobile input: ABBYY Vantage offers a mobile capture that allows capture directly by the end user; additionally, it is also possible to use a shared directory, a mailbox, or even the REST API. The documents and photos are sent to a transaction that can be monitored directly on the server.



Assemble: This activity is responsible for organizing the documents within a transaction. It is possible to create documents according to predefined rules from the captured images and pages. In this use case, police reports and invoices can have more than one page, while the other images and pages form one document each.



Document or image classification: Using an ABBYY Vantage Classification Skill, it is possible to classify each document into an image (car and odometer) or document (driver's license, accident report, and invoice). This classification defines the path to follow in the process for each document.



CV-based image classification: This activity uses a Microsoft Custom Vision API to identify whether the image is a car or an odometer. Again, the result of this classification defines the path to be followed in the process.



Vehicle position identification: In case the image is classified as a car, the Microsoft Custom Vision API is again used to identify the position of the vehicle (front/rear/left/right). The position and confidence rate are returned and saved as document metadata.



Odometer reading: This is the most complex implemented activity because, first, you need to find the odometer in the picture. To do this, a Microsoft Custom Vision API was again used to locate the odometer position in the vehicle's dashboard photo. A Python code handles the return and generates a new image from the coordinates of the region where the odometer was identified. The same Python code sends the image to Vantage to perform the recognition through optical character recognition (OCR). In this use case, the python code was implemented in an Azure Function that receives the image and returns the odometer value. The result is stored as document metadata.



Document text extraction: This activity is responsible for generating the complete OCR of the document, which will serve as the basis for both classification and metadata extraction.



Document classification: Again, using the ABBYY Vantage classification skill, this activity classifies and separates the documents that will be processed by the IDP (invoice and/or identity document) from the documents that will be processed by the LLM (Incident Report). The result of this classification defines the path to be followed in the process.



Document data extraction: Before extracting data from a document, it is necessary to classify it into a specific type that defines which Vantage Document Skill will be used at the extraction step. The process classifies and extracts data from the identity document and/or invoice.



LLM extraction: The basic fields are extracted through the IDP extraction activity. After this extraction, only the description of the incident is sent to the LLM to return a summary and answer whether there were victims or not. The data is stored as document indexes.



Review: After all documents have been processed, it is possible to include a manual validation step (human-in-the-loop) so that if any errors are identified, a human operator can act.



Output: Natively, ABBYY Vantage can export the data in JSON or CSV, and the images in different formats, such as PDF, TIF, or JPEG. Additionally, it is possible to customize an integration for other process management systems or document repositories.

This usage scenario using ABBYY Vantage demonstrates how hyperautomation requires an integrated approach, combining specialized technologies such as intelligent document processing, computer vision, and large language models, among others. Merging these capabilities provides not only a deeper and more comprehensive understanding of data but also enables the automation of workflows in complex organizational environments.

By using these technologies, organizations can automate processes, increase efficiency, reduce errors, and promote smarter, more adaptable automation and a better customer experience. This integrated approach not only drives productivity but also positions businesses to address growing challenges, enabling an agile response to ever-evolving market demands.

ABBYY purpose-built AI for intelligent automation

ABBYY combines innovation and experience to transform data from business-critical documents into intelligent, actionable outcomes in over 200 languages in real time.

We enable more than 10,000 companies globally, including many of the Fortune 500, to drive significant impact where it matters most: customer experience, operational excellence, and competitive advantage. The world's leading enterprises trust ABBYY to get results like 91 percent reduction in invoice processing transaction costs, 99.5 percent accuracy, and 95 percent straight-through processing.

ABBYY is a recognized leader in intelligent document processing by 10 analyst firms just in the past year. For a more detailed analysis of ABBYY's leadership in intelligent document processing solutions, please refer to **Everest Group Intelligent Document Processing Products PEAK Matrix® 2023**.

ABBYY Vantage—Intelligent Document Processing

ABBYY Vantage is the industry's first low-code / no-code intelligent document processing platform, built specifically for the needs of the enterprise. Vantage provides pre-trained AI models that understand each of your document types (structured, semi-structured, and unstructured), enabling straight-through, touchless processing of document data with unprecedented speed and accuracy.

Contact ABBYY today for a demo.