NTT DATA

AI solutions in broadcasting companies

How broadcasting companies can benefit from AI technologies

Pietro Scarpino, Adriano Manfrè, Claudia Lunini

nttdata.com

Index

Introduction

"There is no alternative to digital transformation. Visionary companies will carve out new strategic options for themselves — those that do not adapt, will fail."

Jeff Bezos, Amazon

The digitalization process we are witnessing is profoundly changing business models and is affecting the value of existing goods and services. The resulting digital transformation is creating new needs that require faster and more efficient processes. For this reason, companies across all the industries need a profound transformation in order to accommodate these innovations.

In the broadcasting context, in particular, new technologies have allowed consumers new types of experiences. Viewers in fact, have access to and consume more content than ever, anytime, anywhere. They expect high quality and personalized content from their various providers, along with a different experience across multiple devices. **Broadcasters need to produce multiple versions of the same content at faster rates and lower costs**, while competition for viewers' attention increases for both live content and Video-On-Demand.

Among the technologies used in the broadcasting industry, Artificial Intelligence (AI) is proving to be one of the most effective tool to integrate in several areas and industries. In fact, the AI capabilities have experienced a rapid growth in recent years and the companies have been increasing their interests for these tools accordingly. While companies in many areas have employed the AI in their pipelines, **the AI adoption in the media and broadcasting companies' processes is still at an early stage.**

However, different AI applications in such fields have provided potential results with the consequence of bringing several benefits as well as having a major impact in the companies' value chains.

AI Providers and System Integrators

The companies that have approached the development of AI solutions are of the most diverse and offer always updated and competitive services. Particularly, in the recent years, big cloud vendors have significantly expanded their offering around AI solutions, providing both working solutions (APIs, SaaS products, pre-trained models, etc.) and tools to facilitate the creation of new Machine Learning services. Furthermore, new startups are born continuously developing AI capabilities that cover the most diverse and specific needs.

The use cases enabled by AI are indeed very diverse and often regard specific domains, where ready-to-use AI solutions most likely fails or poorly perform. Therefore, these solutions often need to be enhanced and customized in order to meet the use case requirements. In this scenario, the **system integrator plays a crucial role, combining deep technical knowledge and business vision while working close to the clients and their needs**. NTT Data, thanks to its experience as Trusted Global Innovator, has the expertise to create customized end-to-end products that can exploit the power of the big vendors' tools and the startup innovations enhanced by tailor made solutions that adapt to the clients' needs and deliver a greater value.

Challenges in the Broadcasting

Generating and delivering contents for broadcasting purpose is a complex and time-consuming process, composed of many stages in which creativity and routine bond together.

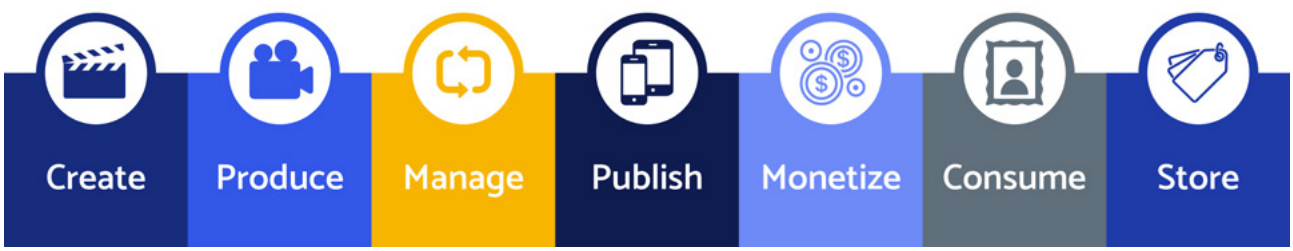
From the content creation phase to the delivery one, passing through the content production and managing, the challenges that broadcasters have to face are multiple.

For instance, in the content management phase the metadata extraction from the videos is required. Such activity consists in tagging each video content with all the information (the more the better) that relate to the video. For instance, in videos, information such as category (e.g. Sport, News, Entertainment, etc.), relevant objects, locations and main topic discussed need to be associated to the content. Another example of relevant information is identifying the people that are speaking. For instance, during the political elections is important to ensure that politicians of different parties have the same amount of time in expressing their thoughts. This activity relies on a group of human operators that, by looking at the content, annotate who is speaking, when, and for how long. This process requires a stressfull mental effort and it is time-consuming.

Another expensive activity in terms of time and human effort is the Video Quality Control (VQC), which is essential to distribute regulatory compliant contents. This procedure consists in a visual analysis, performed frame-by-frame, which aims to determine whether anomalies occur in the content or not. The anomalies, for instance, concern the quality of the video that may be due to old footage even if the content is in high resolution, or the brightness or colors of the frames, while some anomalies may concern the presence or absence of the right subtitles. Another series of anomalies concerns the presence of specific areas of the video where some signs with foreign text appear without the correct translation.

A further costly task regards the generation of replays and highlights. It is hard to imagine watching an event or a live sports broadcast without highlights. All the platforms and the channels deliver sports highlights at an increasing rate. The manual work required creating five or ten-minute recaps of live sports events is significant. In addition, human attention bias has also to be taken into account: humans inevitably miss things.

Finally, for further enhancing customers' experience, the possibility to show relevant information overlaid on the TV represents an extremely appealing feature. To allow that, broadcasting companies have numerous teams, each of them with experiences in different areas. In fact, for instance, while the cinema contents need to show data about actors and characters, the sports events need to manage different



data. For example, in soccer events, statistics on players, and how the teams are moving around the pitch are relevant information. Differently, in motor races, it is important to show different views (e.g. inside the car or in a particular position of the track) as well as the on-going racers' positions in the track. Finally, cooking shows are an important channel to monetize with sponsored objects (e.g. kitchen tools) and, hence, it is important to provide information about them.

The challenges in this industry are certainly still many, but they do not represent the heart of this document.

However, these few examples are sufficient to demonstrate that **AI capabilities could play an active role in overcoming the challenges facing the broadcasting industry**. The AI adoption, in fact, provides not only the automation and acceleration of processes but can also give high value-added improving the effectiveness of such processes and reducing the costs.

AI Technologies for Video Analysis

Teaching Machines to See

Of all of the Artificial Intelligence technologies, **Computer Vision** is one of the most appealing in the field of Broadcasting and it is, in general, among the most promising ones.

Computer Vision is the computer science field that focuses on the replication of the capabilities of the human vision system, enabling computers to see, identify, and process images in a similar way that humans do. It is a multidisciplinary field, which may involve the use of specialized image processing methods and learning algorithms. The variety of tasks that Computer Vision can carry out such as Image Classification, Object Detection, Face recognition, Action recognition, Emotion recognition, Optical Character Recognition (OCR), etc. makes it clear why this field is widely applicable.

In the 1960s Computer Vision started to take shape as a field. It tried to mimic the human vision system combining standard image processing techniques with general learning algorithms. The 2010s saw dramatic progress in image processing gaining popularity among the industry and research communities. This is due to the innovations and results obtained by applying the Deep Learning (DL) techniques to image classification tasks. The recent advances in Artificial Intelligence and in particular in Deep Learning and the maturity reached in the Big Data technologies, allowed Computer Vision to succeed and even to surpass humans performance. As an example, recent **Deep Learning algorithms outperformed humans** by obtaining a smaller error in some tasks related to detecting and labeling objects.

The success of Deep Learning techniques relies on the power of Neural Networks (NN). They consist of a set of biologically inspired algorithms belonging to the learning methods category and nowadays are among the most powerful ones.

Their power lies in a complex structure, made of different layers each of which is composed of different units, called neurons that elaborate the input. The input passes through several hidden layers where it is transformed by different functions until it reaches its final form. Neural Networks are particularly suited to extract automatically complex patterns and relationships that exist within a set of images. In fact, NNs, while trained on a large amount of labeled data, incrementally learn what are the relevant features related to the problem to assess.

One of the most involved Neural Networks in the field of Computer Vision is the **Convolutional Neural Network (CNN)**, which is a particular type of Neural Network that played a central role in increasing the popularity and success of Deep Learning. Their architecture tries to replicate the connectivity pattern and the organization of the neurons of the Visual Cortex in the Human Brain. They are particularly suited to elaborate images and their use requires almost no pre-processing. In fact, while in primitive methods filters were hand-engineered to fulfill the Computer Vision tasks, these networks can automatically learn the set of filters that best suit the task. Particularly, CNNs manage to create the filters so that meaningful information are extrapolated from the images. Examples of such information are the edges and the shapes of the objects, that will be used to gain the ability to distinguish among/between different classes of images.

Empowering Computation with the Cloud

Another important factor that has had a huge impact in the success of Computer Vision is the computational power given by theCloud Technologies. Cloud computing, in fact, allows handling, storing, and computing a huge amount of data in the same place and this in turns allows deep learning models to be properly trained in order to make them effective. Moreover, main cloud providers such as Google, AWS, and Microsoft also dispense powerful cognitive services able to cover several Computer Vision and other AI tasks. These services can be simply used via API calls, even without Artificial Intelligence deep knowledge.

Bringing the Computation Closer to the Sources

The great advance in hardware technologies also played an important role in the diffusion of AI. In fact, Computer Vision and image processing algorithms can be computationally intensive and they often exceed real-time capabilities. Other than Cloud Computing, also modern GPUs allow a wide range of computer vision algorithms to be performed.
Moreover, whenever limitation to standard deployment approaches fail, as for example, network connectivity is not available, or network latency is insufficient, or there are privacy and security concerns Edge Computing comes in handy. New devices, in fact, allow the local computation of AI algorithms on the “edge”, with little or no network connectivity. With Edge computing, images and videos can be brought closer to their sources, so that the latency issues can be decreased, opening new scenarios for the application of the Computer Vision, and security and privacy concerns can be addressed as well.

The Adoption of AI in the Broadcasting Industry

Broadcasters can **leverage AI not only to automate** repetitive and time-consuming tasks currently done by human operators, **but also to support the operators in their activities, enhancing their experience**, making the job easier, less repetitive leaving more room for creativity and cognitive resources.
Automation is not the unique advantage that AI can bring. These technologies in fact can really create value as they have the ability to replicate most of the human decision strategy, in a much shorter time. Moreover, **AI technologies can reduce human errors and can improve the accuracy and the efficiency**.
Artificial Intelligence, and in particular Computer Vision, finds a wide range of applications in Broadcasting companies, bringing benefits in many stages of the digital content chain.

Smart Content Creation

Starting from the creation stage, the first stage of the content chain, AI finds interesting applications supporting operators in the creation process, and automating those tasks in which creativity is not required. This in turns allows human operators to focus on tasks where creativity is required.
The Edge computing finds here a very appealing application: using AI applied on edge to automate camera operations. For instance, automated image capture solutions have been introduced for broadcasters and sports clubs. This solution consists of a robotic video capture system that offers automation, flexibility, and low-light image quality; it autonomously moves the camera and adjusts the zoom and focus to automatically keep the team or the player in the frame.

Smart Content Production

The production phase is very broad and rich in activities where Artificial Intelligence can give a valuable contribution. For instance, the manual generation of captions and subtitles in a video can be very expensive. Conversely, speech-to-text functionalities, together with automatic language analysis and translation enable near real-time automatic subtitles and captions generation in several languages. These procedures, that take humans several hours, can be done instantly by a machine.
In addition, for what concern the metadata extraction and generation Computer Vision techniques are effective. They allow automatically detecting relevant objects in the scene, recognize people, locations, and even classify actions. Computer Vision proved to be also powerful in the monitoring phase, simplifying and fastening the video quality check process.
The generation of replays and highlights can be performed by AI technologies as well. Computer Vision techniques can perform this process in a much shorter time reaching an even better result, since it can simultaneously detect several key scenes near real-time and it doesn't suffer the bias to which the human operator is subject in watching a scene.
These are just few examples that show how the introduction of AI, in particular Computer Vision, can support humans in several tasks in the broadcasting company, and give insights of the potential of these technologies.

NTT DATA Video AI Platform: AI for enhanced broadcasting processes

Thanks to a **close collaboration with the main players in the Broadcasting industry, NTT DATA has developed a strong awareness of their challenges and needs**, which allowed providing winning AI solutions able to support media operators in different domains/activities.

Leveraging R&D initiatives, internal investment, and partnerships with broadcasters NTT DATA ideated and developed **Video AI Platform: a modular solution designed to enhance Broadcasters' processes with Artificial Intelligence**. It consists of a platform able to handle several needs of this industry exploiting Artificial Intelligence technologies; it allows the integration, orchestration, and combination of multiple AI technology providers, including NTT DATA's proprietary technology.

Video AI Platform architecture has been designed to guarantee high flexibility. In fact, it is possible to add new tailor-made capabilities, algorithms and integrate them with the standard cognitive services to satisfy the customer needs, and the integration is effortless. This gives the solution, the power and efficiency of the external services and, at the same time, the flexibility to plug custom models for a full customization.

Video AI is cloud-agnostic and offers the integration of the widest variety of cognitive services, whether they are provided by cloud providers or by startups. This lets to choose the right technology among the multiple available to get the best performance and saving costs. Furthermore, the platform allows including also custom models developed with advanced Deep Learning and Machine Learning techniques necessary to accomplish specific tasks in which is important to consider the domain peculiarity.

Thanks to the modular architecture, the Video AI platform can be fully integrated with clients' pre-existing processes and tools, letting optimization of internal processes with consequent savings in time, reduction of management costs, and minimization of operational risks.

The platform is designed to enable the paradigm of **"human assisted AI"**: human experts help AI become more accurate over time and in turn, AI helps human experts be more productive. Thanks to the user-friendly interface, operators can interact with the platform, personalizing configuration, validate and modify the output through features and tools (dashboard, consoles, reports) having total freedom.

In the Broadcasting value chain, the Video AI Platform is able to serve several processes, enabling the automation of content production, packaging, and distribution. The idea behind the design of Video AI platform, in fact, is to have multiple working modules that can be employed to produce relevant data for a particular use case; it includes different modules each of which is responsible for a specific task.

So far, the modules included support operators in the metadata production and speaker recognition, in the highlight generation, and in the quality control of the video.

Extracting Metadata

Thanks to the metadata production module the Video AI platform allows supporting the operators through the automatic generation of useful information from video contents, as people and objects appearing, audio transcription, and so on. In detail, the module performs the following main tasks: Object recognition, Face Recognition, Scene/Environment recognition, Audio text transcription in different languages, Text summarization and keywords, key-phrase, and topic extraction.
The metadata production module is composed of both computer vision and NLP technologies to handle the audio data as well as the video contents and provide as output the generated metadata. This module offers as a feature the possibility to train the system on new objects or people of interest that can be easily added into the knowledge base, or even more, to personalize the vocabulary of the NLP engine adapting the extraction to a specific domain improving the quality and quantity of metadata according to the context. This gives the users great flexibility and all the customization they need.
Artificial Intelligence, and in particular Computer Vision, finds a wide range of applications in Broadcasting companies, bringing benefits in many stages of the digital content chain.

Recognizing who is speaking

In order to support the operators in identifying the people that are speaking within a video, the Video AI platform has been enriched by providing the module for the “speaker recognition”. This module performs the analysis of the speakers in nearly real-time, and outputs the speaker’s information over the content, leaving the operators in evaluating such outputs with ease. The module is able to uniquely identify in a video/audio, given as input, the segments in which a person is speaking. Specifically, it creates a fingerprint of the voice in order to distinguish a person among the others. To create the fingerprint, a neural network was developed with the goal of mapping people’s voices to unique mathematical representations.

Generating Highlights

The generation of the highlights is an additional feature provided within the NTT Data Video AI platform. Highlight generation is a crucial part of the production stage, in particular for sport related contents and, it is a very expensive task. The highlight generation module consists of a system that automatically creates highlights of video according to the user’s preferences. The module employs Deep Learning techniques to create a scene-importance estimation model that estimates the importance of each scene within a video. In order to understand the importance of a scene, the algorithm considers several features: emotions of people appearing in the video, actions in the scene, as well as noise in the audio (cheers for example) and the information of the score, in particular for live sports event.

Checking for Anomalies

Another crucial activity in the broadcasting process is the Video Quality Control that can also be performed by operators using the NTT Data Video AI platform. The platform contains, in fact, the Augmented Video Quality Control (AVQC) module, which is able to handle potential anomalies in the video content exploiting different AI algorithms. Such module represents a supporting tool for the quality control operators during the VQC process. The module leverages several custom AI capabilities, from statistical analysis to Deep Neural Networks, as well as integrating cognitive services.
Since the visual anomalies (e.g. color inconsistencies, ancient patterns, etc.) in the video depends on the context of the scene, the AVQC module has been designed and developed with several custom deep neural network to overcome the lacks of cognitive services provided.
The custom neural networks developed take inspiration from the State-of-the-art in DL. In particular, these DNN can extract features from the video’s frames and analyze the sequence of features in order to detect a potential anomaly according to the context. For some anomalies, a hierarchical architecture composed of multiple neural networks has been designed. Each neural

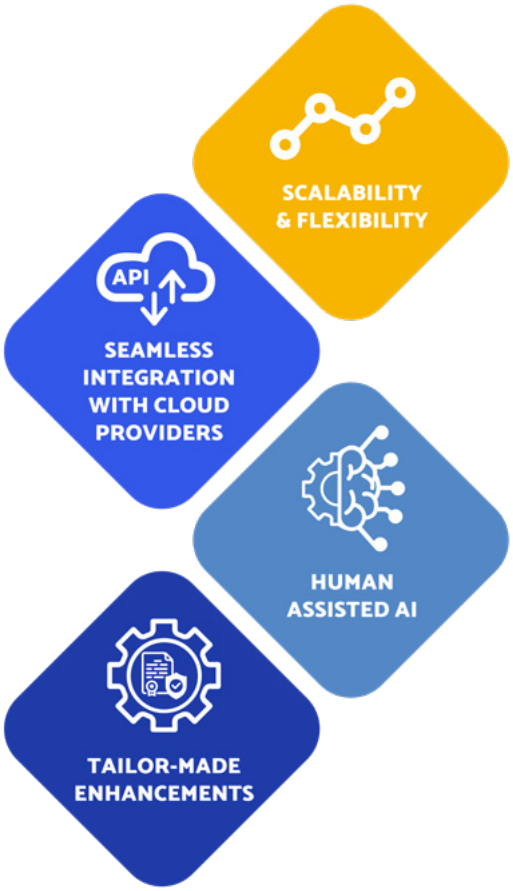
network contributes, extracting intermediate information, to accomplish/reach the final goal. For example, in certain anomalies, the intelligent engine needs to determine which are the areas in the video that the viewer is focused on, correlating the information of consecutive frames. To do this, a specific Deep Neural Network, in the intermediate step, can determine the relevant area within which the last network will detect the anomaly.

Enhanced TV

The Video AI platform allows the extraction of many information such as data about actors and characters appearing in a video, statistics about players and how the teams are moving around the pitch, for what concerns sports events, and other information of interest for the consumers. In fact, data from the metadata production module about objects, scenes and actors, can be combined with the speaker recognition module that identifies who is speaking in entertainment content, so that users have all this information at glance. Another example is regarding the sports events, in which the highlights generator module already sums up the most important scenes of the event that are brought to the customer's attention.

The modular structure of the platform allows adding new blocks for accomplishing specific tasks. In particular, new modules as for example the data visualization for sports events (e.g. soccer teams statistics, racers positions, etc.), can be easily included to enrich the content to display, bringing new incredible viewing experience to the final user.

Given the platform capabilities, it's easy to understand the benefits deriving from its use: it minimizes the human effort required to carry out the activities, increases the video contents analyzed and the efficacy of the services, and at the same time, reducing the costs and the time.



NTT DATA has providing winning AI solutions able to support media operators in different domains/activities.

Augmented Quality Control

Metadata Generator

Sport/News Higlights

Voice Speaker Recognition

Explicit Content Alerting

Real Time Advertising

Conclusions

The Artificial Intelligence in media and broadcasting is proving to be a valuable ally and is delivering tangible outcomes. Companies that have adopted AI-driven solutions have gained benefits in terms of increasing process efficiency, optimizing the internal workforce, and reducing costs. Moreover, AI capabilities are growing at an incredible rate and the possibilities of applying AI to this industry are still many and are likely to increase even more.

Nowadays, the AI market is very rich and competitive and offers many cloud computing services, algorithms ready-to-use, and higher computational power that enable everyone to exploit AI capabilities. However, it is not always simple to choose among the several options already available, and this is not only a technology matter. There could be many other external and internal factors playing a key role in the definition of an AI solution. For instance, the introduction of an intelligent engine into the pre-existing client infrastructure is not always straightforward and it can affect the processes in several steps. Moreover, it requires the awareness and knowledge of AI, the trust in this new technology, and last but not least the people's expectations and perception of the AI introduction in their daily lives.

The NTT DATA extensive experience in the broadcasting industry together with the deep expertise in AI technologies can support customers in navigating correctly towards these technologies, and allow creating valuable products that respond to the client's challenges and better suits clients' needs, bringing values by enhancing the processes and optimizing the workforce allocation.

The NTT DATA Video AI Platform, in fact, already constitutes a support tool of great impact for broadcast operators, facilitating expensive tasks like the metadata extraction, the highlights generation, and the VQC process. Nevertheless, the flexibility with which it has been designed guarantees the possibility of an evolution that aims to cover more and more needs, making the platform a product in continuous transformation. Such transformation would be carried out by broadcasters themselves based on their needs by teaming up with the NTT DATA's experts that provide support, experience, and expertise in such area.

The NTT DATA Video AI Platform, constitutes a support tool of great impact for broadcast operators, facilitating expensive tasks

Contact person

Pietro Scarpino
Head of IoT, VR & AI Service Line

NTT DATA Italy

Adriano Manfrè
AI Team Leader

NTT DATA Italy

Claudia Lunini
Data Scientist

NTT DATA Italy

NTT DATA – a part of NTT Group – is a trusted global innovator of IT and business services headquartered in Tokyo. We help clients transform through consulting, industry solutions, business process services, IT modernization and managed services. NTT DATA enables clients, as well as society, to move confidently into the digital future. We are committed to our clients' long-term success and combine global reach with local client attention to serve them in over 50 countries. Visit us at nttdata.com

NTT DATA