NTT DaTa

How media companies can create an Enhanced TV experience

Providing better interactive experience and engagement by moving audience from passive into active watcher

Ruggero Di Benedetto, Fabio Pavanetto.

nttdata.com

**Advisory Services
Together, we innovate.**

Index

What is an Enhanced TV experience?

If you ask someone "where do you watch TV?", most of the time the answer is "on the sofa". TV experience is often associated with the words "lean back experience": you sink into the sofa, you switch on the TV and switch off your brain for enjoying the show.

But it hasn't always been the case. The first experiment of interactive film is dated 1967, titled "Kinoautomat", and in some points of the movie an actor invited the viewers to make a choice between two alternatives story path using the remote control.

Today, with the new technological infrastructures, it is possible to develop an ever-greater interaction and involvement with the viewers. Thanks to the advent of the internet, TV has become connected and now it can use all the services provided by the web to enhance the viewers' experience, by adding different levels of interactivity.

Enhanced TV represents the intersection of media content and new technologies. Thanks to advanced content creation techniques and technologies, users can enjoy new formats in which they can gain additional information, personalize their experience or interact by influencing what they are watching.

This is not just a matter of technology, but it is also related to the content. Currently, most of the content is created by teams of professionals with considerable time and costs. To create great immersive experiences, capable of transforming the viewer into an active actor of the scene, it is necessary to exploit new mechanisms for generating content. A great help comes from artificial intelligence technologies, which can tremendously increase the quantity and quality of generated content.

This document aims to provide an overview of the current situation of interactive programs and services with a focus on the trends that we believe are the most significant, in particular in the fields of cinema, entertainment and sport. Furthermore, we are going to provide insights into some main aspects related to develop and distribution of interactive contents.

Enhanced TV represents the intersection of media content and new technologies.

Interactive experiences trends in the media and entertainment industry

Movies

The world of interactive films and TV series is experiencing a strong evolution especially among network video distributors and broadcasters, where a strong growth for this kind of contents is expected in the near future. They can be considered as scenes connected in a non-linear sequence through key points linked to other videos, or texts, and / or images.

While watching, viewers can interact with the content to have in-depth information such as additional content, actor biographies, trivia (e.g. Amazon x-ray) and/or are able to create or join "communities" to exchange opinions while they are watching in sync (e.g. NetflixParty or Disney+ GroupWatch).

Another example is spectators shaping the destiny of the actors and the plot through the selection of several options available in the different moments of the story. The most famous example is Netflix Black Mirror's Bandersnatch where this mechanism allows to obtain different story lines and endings based on the chosen combinations.

The data collected through the interaction could be used for marketing purposes in order to obtain more information about customer's habits and preferences such as, what food they choose for breakfast or what type of cars and brands they prefer.

These examples suggest that the era of "interactive storytelling" will play a significant role in the future of home and cinema entertainment.

TV Shows and documentaries

There are different ways in which an entertainment program can involve and engage viewers:
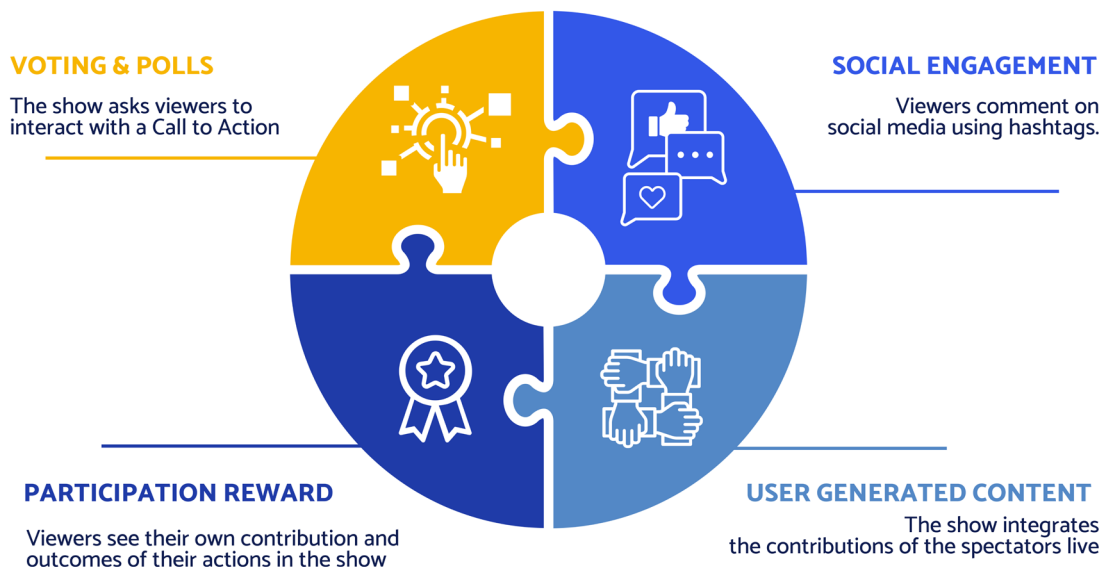
- **Voting and Polls:** the public participates during the show through the use of apps/web sites, influencing the outcome of a specific phase of the show. For example in the case of *X-Factor* the outcome of the performance of each competitor is decreed by the public voter during the live show.

- **Social Engagement:** entice the public using hashtags and comments on social networks and, once a target moment is reached, the broadcaster can unlock additional content encouraging greater participation.

- **User Generated Content:** as a way to further boost participation, a broadcaster/media company can even involve viewers to produce user generated content (UGC) to be embedded in the show. In 2017 Fox Channel asked viewers to send memes made with Simpsons characters using the #memefest hashtag, which were then presented in channel before the advertising space.

- **Participation rewards:** the broadcaster/media company can give the chance to win prizes through the participation to a giveaway; in this way loyalty increases along with the viewer involvement in the show.

In the same way as TV Shows, news and documentaries have faced the same "upgrade". A news network can now create interactive news contents such as news of the day, contextual information, or interview transcripts. News of the day consist of news headlines, while contextual information are maps or statistics helping viewers to situate or understand the impact of some breakthrough events.

In interactive documentaries, just as movies, viewers can choose what the main character must perform (actions, comments, etc.). After each choice they can see immediately its effects, e.g. when encountering a wild animal or when a river must be crossed, and so on.

VOTING & POLLS
The show asks viewers to interact with a Call to Action

SOCIAL ENGAGEMENT
Viewers comment on social media using hashtags.

PARTICIPATION REWARD
Viewers see their own contribution and outcomes of their actions in the show

USER GENERATED CONTENT
The show integrates the contributions of the spectators live

Sport

In the sports world, the concept of interactivity has become increasingly important in recent years through the use of similar fan engagement techniques (mechanisms of request, participation, reward, etc.).

New technologies allow viewers to watch an event (for example a baseball game) from multiple points of view, to choose the camera from which they want to see the actions, and to enjoy different perspectives in real time. This has been even more important during current pandemic times, when stadiums were empty and fans couldn't enjoy the live experience.

Fans can even directly affect or direct the performances of the athletes and the results of the game. Recently, Formula-e introduced a new mechanism to engage with millennials, allowing them to vote for their favorite driver before the race through social media or a dedicated app, in order to give driver extra electric power boost to be used during the live race for five seconds.

Transforming a passive audience into active fans is a trend that was premiered by the E-sports world. Accelerated by the lockdown period, this sector has seen an enormous increase in terms both of revenues, growing up to $1.1 billion in 2020, and loyal fans, with an audience reaching 500 million users. This phenomen was also due to leverage user-generated content able to engage viewers in an entirely different way and introduce new and innovative revenue streams.

Video Shopping

The next level of shopping is Live Commerce, a new way to bring interactive experience into e-commerce. Already widely spread in China, for Live Commerce we mean one-to-many live video shopping or enhanced video content in which viewers can discover a featured product, add directly to the cart and then to buy it through interactive overlays integrated into the web/app player.

Through live streaming influencers can make personalized demonstration of sponsored products and they can respond in real-time to viewers' questions, manage their reactions and further stimulate their engagement using interaction mechanisms.

Through enhanced video content it is possible to use technologies that identify and track objects, clothes, or other stuff within the entire clip. Viewers can click on what catches their attention, find out more information and finalize a potential purchase. This allows to transform each video into monetizable content thanks to product tracking and the activation of referral mechanisms.

Thanks to these dynamics it is possible to create online auctions or massive shopping events. This digital experience makes the online shopping activity more interactive than ever and offers the opportunity for viewers to purchase products as they are shown on screen without interrupting their viewing experience.

While watching, viewers can interact with the content.

Technical challenges to enable an Enhanced TV experience

To leverage the business advantages provided by interactive contents, it is necessary to pay attention to the technologies enabling this kind of solution.

Based on our experience, we identified four technical challenges that need to be addressed to deliver a proper Enhanced TV experience.

Metadata generation

One of the key pillars to enable an Enhanced TV concept is to leverage metadata. They can be defined as a set of information describing other data and each digital object, even the smallest, has them. For example, for cinema content, metadata can be information describing cast or soundtracks while in sports statistics and additional insights of the game.

The biggest challenge in generating metadata is time. Someone can argue that time can always be translated into cost and therefore a solution is just a matter of budget allocation. This is not always true, especially in a media context where the content source can require to work on a real time basis rather than on pre-processing production.

In the latter case the metadata can be generated and collected in advance and made available to the viewer. It is necessary to know the detailed schedule of the contents that will be broadcasted in the next days and to process them in advance in order to refine the metadata generated with editorial support.

METADATA GENERATION **LOW LATENCY DISTRIBUTION** **PLAYER SYNCHRONIZATION** **CONTENT OVRELAY**
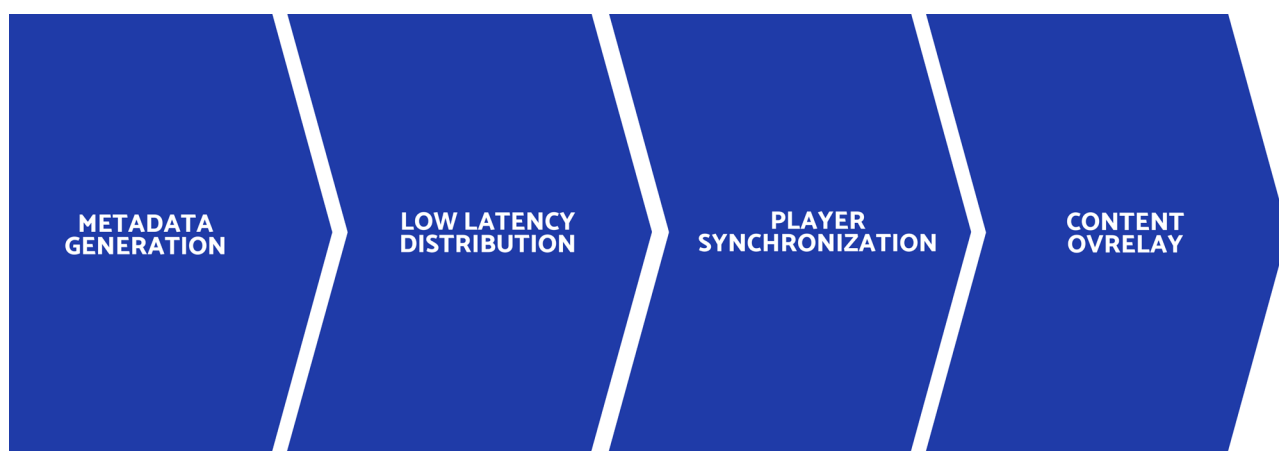
In live-events, instead, the generation of metadata must take place in real time: in this case the difficulty lies in the need to obtain quickly, with a high degree of accuracy and without human support, contextualized data describing the live content to be delivered simultaneously with the video stream

Low latency distribution

In sports low-latency has always been a must-have feature independently from the media (being satellite, cable or streaming), as nobody likes the spoiler effect of listening to your neighbors when they are screaming for a score you haven't seen yet.

With the evolution of expectations and the diffusion of interactive live contents, it becomes even more essential to manage latency in order to give the viewer the desired experience.

This situation is improving thanks to the adoption of a standard streaming format enabling lower latency and the spread of high-speed broadband technologies such as 5G.

In interactive content low latency is at the heart of the deep interaction between spectators and content. Several platforms leverage low-latency streams to embed in video contents mechanisms that allow the spectators to influence what they are watching through an instant feedback, creating new forms of social involvement.

It is clear that having the right infrastructure, exploiting new technologies and delivering a low-latency stream is essential to create new meaningful interactive experiences.

Player synchronization

Showing additional data on the screen and leveraging the interactivity mechanisms adds a further level of complexity to the classic audio / video distribution systems, due to the need of synchronizing this stream with the data shown on top. Synchronization becomes even more challenging when the content source is from a

live event due to the need to provide real-time generated metadata

Being able to offer synchronized stream on all devices enables media companies to create new content that leverages interaction and sociability, allowing the viewers to obtain additional information perfectly synchronized with the content being played, avoiding accessing additional information in advance.

Content overlay

For Content Overlay we mean the feature to add new/additional information over content already present on the screen. Interactivity can be perceived as valuable only if it is a fluid experience rather than a clunky or static one.

The insertion of the overlays is a job that requires the development of several components capable of satisfying all the needs of the viewer, with a huge impact on costs and on the ability to adapt to new demands from the public quickly.

Many companies that deliver interactive applications and contents have developed overlays management systems that help them create custom overlays fast and reduce development costs according to the principle of "build one, deliver many".

Enhanced TV Case Study

NTT DATA has been supporting media clients into enhancing their customer experience and product quality even when facing the technical constraints described in the previous chapters.

One of our client, a leading broadcaster, was looking to strengthen the quality of its cinema/entertainment channels by leveraging interactivity.

As trusted partner, NTT DATA proposed to launch a proof of concept for designing a future-proof Enhanced TV experience.

This Enhanced TV concept not only creates a new way of enjoying movies or TV series by letting the viewers discover more relevant and personalized content, but it optimizes also the

generation of metadata required to enable this experience, leveraging cognitive services to reduce manual operational activities.

After running several design thinking workshops, to ensure that overall concept would be built with "the viewer eyes" and in line with the client brand values, the solution blueprint was composed by these key features:

- Recognize the actors and provide their biographical information and filmography;

- Recognize the soundtracks, provide the details of the track and allow the viewer to go to the moment they were played;

- Recognize the places where the scenes are set and provide further details and description;

- Provide suggestions and correlations with the movie, the cast and the locations identified;

- Provide additional content related to the movies being viewed.

Additional contents

Correlated

Actors

Soundtrack

Location

Batman *directed by* Tim Burton — 1989 Action PG12

00:20:35 02:15:00

Michael Keaton — *Batman*
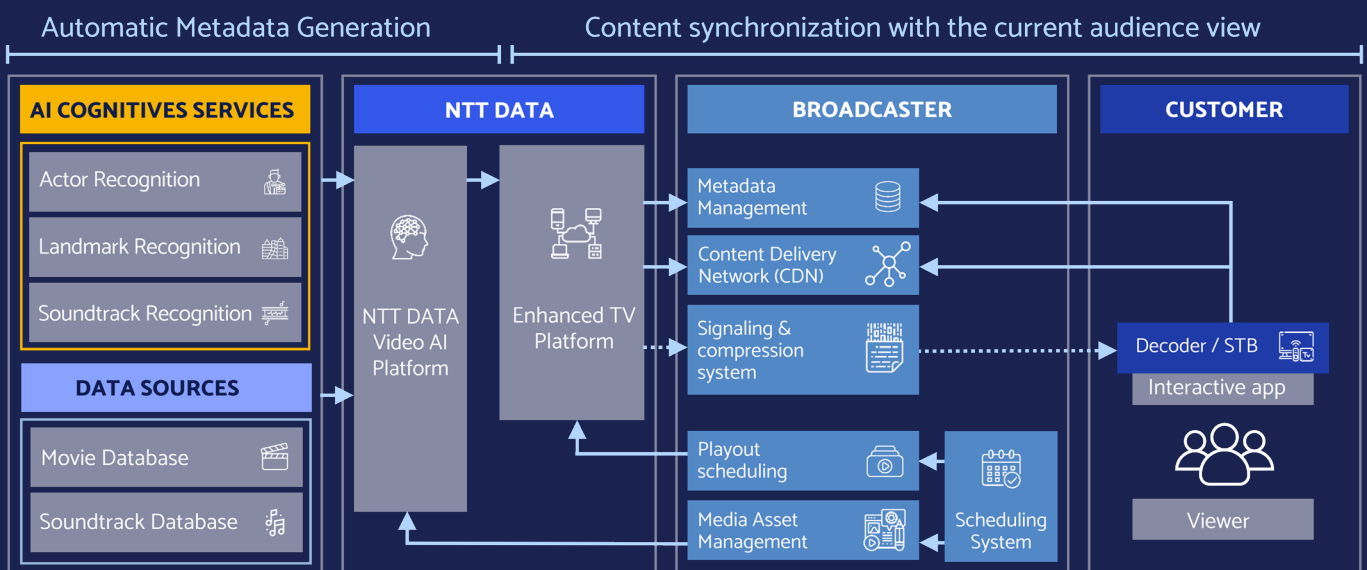
Jack Nicholson — *Jack Napier*

The Future — *Prince*

Bedfordshire

SNYCH WITH THE APP FOR A MORE IMMERSIVE EXPERIENCE

In the following paragraphs we will describe the two most important challenges faced during the project: the automatic generation of metadata and their synchronization with the content being played on the viewer's smart TV decoder (set top box).

Automatic Metadata Generation

Content synchronization with the current audience view

AI COGNITIVES SERVICES
- Actor Recognition
- Landmark Recognition
- Soundtrack Recognition

DATA SOURCES
- Movie Database
- Soundtrack Database

NTT DATA
NTT DATA Video AI Platform

Enhanced TV Platform

BROADCASTER
- Metadata Management
- Content Delivery Network (CDN)
- Signaling & compression system
- Playout scheduling
- Media Asset Management

Scheduling System

CUSTOMER
Decoder / STB
Interactive app

Viewer

Automatic Metadata Generation

One of the key requirements of the concept was correctly identify the frame in which an actor appears or the minute in which a specific music of the soundtrack is played, to provide real time information to spectators about the scene being viewed at that moment.

These tasks can be considered trivial from an editorial/human point of view, but they can be quite challenging for a machine. For example, considering the actor recognition, take a classic movie like the 1989 "Batman": Michael Keaton and Jack Nicholson appear several times with or without wearing their mask, sometimes they even aren't on the screen at all, and the only element in the scene is their voices.

Recognizing actors on the scene and generating metadata automatically has to combine several video and audio analysis services, all running at the same time, in order to obtain a high-quality output in a short time frame with a high degree of accuracy, to avoid the continuous supervision of an editorial operator.
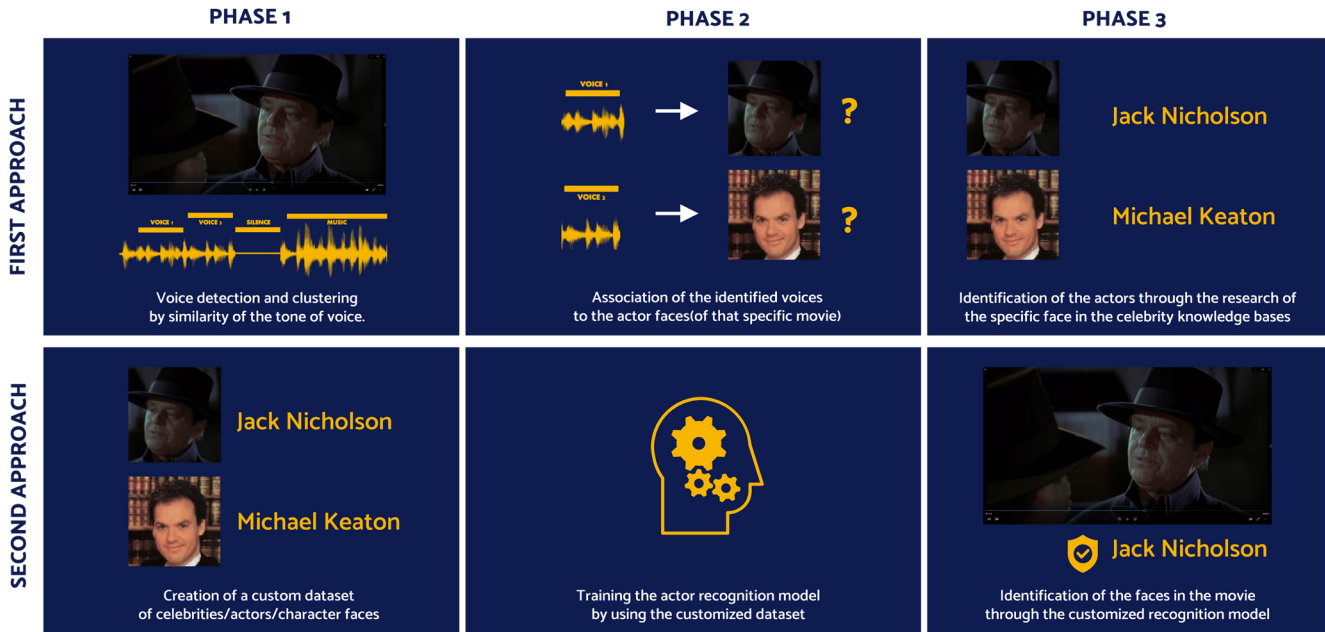
At the heart of this solution there is NTT DATA Video AI Platform, a modular and cloud-agnostic solution, designed to improve content supply chain workflows through artificial intelligence. The platform allows to easily orchestrate several cognitive services with the aim of obtaining the desired metadata automatically.

Content synchronization with the current audience view

Ensuring that each viewer sees additional content synchronized, frame by frame, with the video stream being played is essential to deliver the desired user experience.

Generally, a broadcasting Media Asset Management manages different versions of each film or TV series episode to be aired at different schedules, so for example the prime time version differs from a late night one to include ads with different length or to cut out scenes with explicit contents. This has an impact on the synchronization of metadata with watcher view, because each content version has a different duration.

For this reason, the solution required to merge generated metadata with the playout scheduling data, in order to continuously display additional information with the correct timing synchronization. These data can be collected by the application through the development of specific API or via Content Delivery Network (CDN) , in order to cope with devices concurrency and potential low latency issues.

Device capabilities were also a further challenge for content synchronization. Each viewer can always customize his own video stream using trick mode features like pause, rewind playback or go forward. To allow the smart TV decoder / set-top-box to display contextual metadata for every viewer, the solution required to add an off-set/fingerprint to the video stream in order to calculate the delay between the frame being played versus the video stream actually aired.

This solution allowed to update every single video stream moment by moment during content playback and to bring contextualized cast, soundtrack and landmark information as soon as the viewer tunes in to the channel of interest.

Conclusion

"Enhanced TV" is no longer the future, but the present of the media industry. Offering to spectators interactive content is the key to transform them from passive into active actors, to increase engagement and viewing experience.

Choosing which interactive experience to offer, thinking about the goals and the strategies of the business, is the first step to take. Then it is necessary to design the right technological infrastructure to support this plan. There are several platforms that offer end-to-end solutions for interactivity management; however it is not always easy to choose the one that best suits the different use cases. Furthermore, introducing new components into existing architectures, especially in a broadcasting/media ecosystem, can impact the content delivery process on multiple fronts.

In NTT DATA we praise extensive experience in the media industry, strong technological skills and solutions ready to be tailored for every need. We build valuable partnerships with customers and we support them, step by step, along the creation of these new experiences, with the aim of deliver better products and contribute to the development of our clients' business.

Offering to spectators interactive content is the key to transform them from passive into active actors, to increase engagement and viewing experience.

Key Takeaway

1 Enhanced TV represents the intersection of media content and new technologies that to develop an ever-greater interaction and involvement with the viewers thanks to new formats in which users can gain additional information, personalize their experience or interact by influencing what they are watching.

2 To leverage the business advantages provided by interactive contents, it is necessary to pay attention to 4 technical challenges: metadata generation, low latency, player synchronization, content overlay.

3 NTT DATA developed an Enhanced TV solution for cinema content that recognizes actors and soundtracks and provides additional information, recognizes the places where the scenes are set and provides further details, suggestions and additional content related to the movie.

4 Thanks to NTT DATA's Video AI Platform the Enhanced TV solution is able to recognize the actor and the places on screen or the soundtrack being played, generating metadata automatically with a high level of accuracy, reducing the manual activity of the content editors and increasing the quality of the final product offered to the viewer.

Ruggero Di Benedetto
Director

NTT DATA Italy

Fabio Pavanetto
Manager

NTT DATA Italy

NTT DaTa