Artificial intelligence applications in media archives

Virginia Bazán-Gil

Virginia Bazán-Gil

https://orcid.org/0000-0003-4920-2212

Radiotelevisión Española. Fondo Documental RTVE

Universidad Carlos III de Madrid
Facultad de Humanidades, Comunicación y Documentación
Departamento de Biblioteconomía y Documentación
Madrid, 128. 28903 Getafe (Madrid), Spain
virginia.bazan@rtve.es

Abstract

The aim of this paper is to present an international overview of the use of artificial intelligence in the context of media archives in broadcasters, preservation institutions and press agencies, through a comprehensive analysis of sources primarily focusing on case studies presented at international conferences and seminars, together with the results of the survey on the use of artificial intelligence conducted by *FIAT/IFTA*. Once the most commonly used technologies have been defined and we have identified the stages of the production workflow in which they are used, we will discuss the specific applications of these technologies in television archives, audiovisual heritage preservation organisations, press agencies and innovation projects where technology vendors and media companies collaborate. Finally, we will deal with the challenges related to the implementation of AI in media archives, the need for datasets in the development of language models, and the relevance of a sensible use of technology.

Keywords

Artificial intelligence; AI; Media archives; Television archives; Audiovisual archives; Audiovisual documentation; Speech technologies; Natural language processing; Metadata; Media; Radio; Television; *FIAT/IFTA*.

1. Introduction

In recent years, artificial intelligence (AI) has made its way into most areas of society, to the point of dominating the public debate given the popularity achieved by generative AI systems such as *ChatGPT* (*Open AI*, 2022) or *Dall-E* (*Open AI*, 2023). This is also true for the media industry, which has adopted these technologies in all the phases of the journalistic process, from information gathering to automated content production and distribution, as well as in the media's relationship with their audience (**Sánchez-García** *et al.*, 2023).

In the field of content production, AI enables the analysis of massive data in order to understand relevant news events that generate large volumes of information, such as electoral processes or sports events like the Olympic Games. Although at a slow pace, automated news generation is starting to permeate into newsrooms, outlining a future where professionals will be adding value to information previously generated by algorithms from structured massive data. Additionally, in the realm of production, AI has become an essential element in fact-checking and deepfake detection.

Regarding content distribution, AI allows for alerts and recommendations settings based on user profiles, as well as for automatically generated subtitles and translations. These elements are essential in public media to guarantee that information is made accessible to different groups, such as the hearing impaired. Automatically processed subtitles and informative texts enable the detection of protagonists and keywords that can be used as tags. This improves content discovery and visibility, and attracts a larger number of users.

2. Objectives and methodology

The main objective of this work is to provide an international overview of the use of AI in the context of media archives within television companies, audiovisual heritage preservation organizations and news agencies. It also presents innovative projects that bring together technology providers and media archives to delve into the application of these technologies.

To this effect, a comprehensive analysis of sources, mainly using case studies presented at international conferences such as the *International Federation of Television Archives* (*FIAT/IFTA*) or the seminars and sessions of the *European Broadcasting Union* (*EBU/UER*) work groups, has been carried out from 2013 to the present. These sources of information are only accessible to industry professionals, since the *EBU/UER* seminars and reports generated by the various groups are open to members only, while *FIAT/IFTA* typically imposes a one-year embargo on the content of its annual conferences and seminars. This overview is complemented by the results of a survey on the use of AI conducted by *FIAT/IFTA* in April 2023, whose objective was to understand how media archives are utilizing AI. The initial results were presented at the *Media Management Commission* (*MMC*) seminar of *FIAT/IFTA* last May 2023.

3. An overview of the present situation

According to an investigation conducted in 2022 by *Radiotelevisione Italiana* (*RAI*) as part of the *AI4Media* project, the use of AI in the media is no longer a new trend, although it is still far from being widely adopted industrial practice (**Bruccoleri** *et al.*, 2022; *AI4Media*, 2023), and it is still uncertain when fully operational and high-quality functionalities will be available. Moreover, some essential tasks are still not well covered by current applications, which implies the need for new ones to be developed. According to this report by *RAI*, these technologies have enormous potential to support the value chain of media organizations, who would be able to enhance the quality and creativity of their work significantly without replacing human labor. Lastly, trustworthiness is a crucial factor in the application of artificial intelligence the media industry, making it essential that the tools developed respect user privacy and comply with data protection regulations.

In the field of media archives, the biennial seminars of the *Media Management Commission* (*MMC*) of *FIAT/IFTA* are credited as a benchmarking framework for the exchange of ideas and knowledge about archive management, media, metadata, and technological advancements (**Green**; **Gupta**, 2019).

- In 2007, in Vienna (Austria), the initial technological advances in automated cataloging, scene detection, face and object recognition, automatic character recognition, and automatic subtitle generation were first presented.
- In 2013, in Hilversum (The Netherlands), the focus was on AI and automatic annotation, and media archives started to question the future role of catalogers, a debate that remains relevant to this day (*FIAT/IFTA Media Management Commission*, 2013).
- In 2017, in Lugano (Switzerland), the first pilot tests on television archives, performed with second generation MAM's (media asset management), were analyzed. Once again, the future role of documentalists was discussed, this time in terms of their active involvement in model generation and results supervision (*FIAT/IFTA Media Management Commission*, 2017).
- In 2019, in Stockholm (Sweden), the 20th anniversary of the *MMC* seminars was commemorated, highlighting the adaptability of a sector that keeps asking similar questions in a constantly changing technological context (*FIAT/IFTA Media Management Commission*, 2019).
- In 2023, in Dublin (Ireland), with numerous archives using AI solutions in production, the debate revolved around the undeniable need for coexistence between algorithms and humans (*FIAT/IFTA Media Management Commission*, 2023).

Currently, AI is being applied in the media both in the production archive and the deep archive. In the production archive, it is used for feed and raw material analysis so as to facilitate the immediate retrieval and usage of footage, particularly in newsrooms. In the deep archive, solutions focus on the retrieval of collections with insufficient description levels to ensure their reusability.

Each institution may choose different solutions as part of their innovation projects, proofs of concept, or other projects with a clear timeframe. It is noteworthy that AI projects in archives are developed by interdisciplinary teams involving different areas within the organisation. Additionally, archival material holds essential value in the development of solutions applied to websites or video-on-demand platforms, sometimes without a direct return to the archive itself, as we will see in some of the use cases presented below.

4. Which technologies are we talking about?

Before delving into specific applications, it is important to define the technologies we are referring to when it comes to AI applied to media archives. Let us take a closer look at them:

- Speech and audio technologies: This term refers to a set of technologies that enable automatic speech recognition and its transcription into text, including language recognition and speaker identification, as well as the detection of certain traits connected to speech, such as gender, age, or emotional state. These technologies also allow for the analysis of the acoustic environment, including speech detection, music, and silence.

- Natural language processing: This concept encompasses techniques that enable the understanding of text structure and meaning. Through its application, it is possible to detect named entities, identify keywords or automatically classify content in texts. These technologies are also used for text generation, summaries, etc. through generative AI.

> In newsrooms, AI enables immediate content retrieval during production. In the archive, collections with very low levels of cataloging are made accessible

- Computer vision: This is the branch of AI that enables systems to extract meaningful information from digital images or videos. In the context of media archives, it is primarily applied to facial and identity recognition, logo and object recognition, subtitle recognition (optical character recognition, OCR), scene and shot segmentation, image summaries and automatic content generation.

These technologies are being implemented in both public and private media organisations, as well as audiovisual heritage preservation institutions and news agencies. In recent years, driven by the *Next-Generation European Funds*, numerous innovation projects have involved television networks that contribute data, knowledge, and expertise. This creates a synergy with the industry, seeking a better understanding of the market to develop products for a sector undergoing significant transformation which is trying to outline its future.

5. Three key actors: Preservation institutions, broadcasters and news agencies

There are three fundamental actors involved in AI projects for the management, preservation, and exploitation of audiovisual collections. Let us now take a closer look at each of them.

5.1. Preservation institutions

In Europe, two organizations are leading the way in audiovisual heritage preservation: the *Institut national de l'audiovisuel* (*INA*) in France and the *Netherlands Institute for Sound & Vision* (*NISV*).

5.1.1. The *Institut national de l'audiovisuel*

The *Institut national de l'audiovisuel*, commonly known as *INA*, is a public institution established in 1975 with the aim of preserving the French audiovisual heritage, creating content, conducting research, and transferring knowledge in the audiovisual and digital fields (*INA*, 2023a). Ina is responsible for the legal deposit of audiovisual materials and websites in France. As part of its mission, it promotes and facilitates access to the collections it preserves for professionals, researchers and the general public. The *INA* collections include the production of 179 television channels, radio stations, websites, and social media accounts (*INA*, 2023b). From a practical standpoint, this entails handling large amounts of heterogeneous metadata from various sources, which can sometimes be inaccurate or even nonexistent. In this context, AI emerges as a suitable tool to improve the description and ensure the accessibility of these contents. In this regard, several innovation projects have been developed in recent years, such as *NOA* and *Trombinos*.

The *NOA* project applies computer vision and natural language processing techniques to segmenting the broadcasts of a television channel into programs, breaking these news programs into individual news stories and identifying the topics discussed through the analysis of captions (**Martín**; **Segura**, 2021; **Couteux**; **Segura**, 2023). computer vision is used for the recognition of title sequences, logos, presenters and credits, thus enabling the identification of the beginning and end of each program and associating it with a title sequence. The presence or absence of the presenter on the screen facilitates the segmentation of the programs into different news stories whereas subtitle analysis helps identify the main topics covered in the news. All the automatically generated information is subsequently manually validated by a professional.

Trombinos is a facial recognition project based on an *IBM* model developed by *INA*. The algorithm has been trained on 62 million faces corresponding to 70,000 individuals. These images have been obtained from both television programs and images retrieved through internet search engines. The content is processed using the model. The results are then returned with a certain level of accuracy and validated either manually or automatically. Each recognized person is associated with an authority record that includes links to external data sources such as *DBpedia*. *INA* is currently working on a less biased model in order to achieve better gender and racial representation (**Petit**, 2022).

In addition, AI-based solutions have played a significant role in the development of data.ina.fr, a portal aimed at promoting knowledge of the Ina collections through data analytics (**Roche-Dioré**, 2023), where *Trombinos* and *INA Speech Segmenter* provide facial recognition and audio segmentation, *Vocapia* performs speech-to-text transcription and *Textrazor* handles entity recognition. The process of analyzing contents massively through the AI platform has generated large amounts of data that can enable the development of subsequent studies related to the media.

5.1.2. The *Netherlands Institute for Sound and Vision*

The *Netherlands Institute for Sound and Vision* is the institution responsible for preserving the heritage of public media in the Netherlands (*Netherlands Institute for Sound and Vision*, 2023) and making it available to society as a whole. *NISV* early on embraced the use of automatic metadata generation and their integration with external data sources in order to enable the exploitation of its collections. In 2012, *NISV* began using automatic speech-to-text (S2T) transcription te-

chniques, but with a high incidence of errors. Simultaneously, they applied Named Entity Recognition (NERD) solutions, a practice abandoned in 2019 since it did not meet their needs in a real production environment. Also in 2012, *NISV* started developing voice and facial recognition to enable content tagging (**Manders**, 2019).

The current facial recognition mode used at *NISV* relies on an onomastic thesaurus which only includes public personalities in accordance with a well-defined privacy policy that respects the *General data protection regulation* (*GDPR*) (**Manders**, 2022). The project team defined the expected accuracy levels from the beginning, so only faces identified with an accuracy level above 90% are ingested into the system. Currently, the system has reached a 95% accuracy level, it recognizes 3 out of 4 individuals and is capable of tagging up to 50% of the faces that appear in a television program. Despite the good results, there are concerns about its scalability, its limitations related to the use of the thesaurus as an essential element, the bias that this implies, the difficulty in identifying emerging personalities and the lack of facial models in the dataset, which can lead to false identifications.

Despite adopting all these technologies and having a strong innovation and development area, in 2022, the majority of the metadata managed by *NISV* had not been automatically produced (**Manders**, 2022; **Manders**; **Wigham**, 2021).

5.2. Broadcasters

Broadcasters face constant challenges since they have to deal with the strain of immediate production and the urge for a large volume of content from different sources to be made accessible and also ingested into their production systems on a daily basis. European public televisions have gathered and preserved content from different sources dating back to the 1960s following changing cataloging policies. For them, AI represents an opportunity to improve efficiency, increase the reuse of archival assets and avoid repetitive tasks. Let us take a look at some examples.

5.2.1. *Yle*

Yleisradio Oy (Yle), the Finnish public broadcaster, has been a pioneer in testing AI-based solutions (**Selkälä**, 2017). The application of automatic techniques for metadata generation was considered a way to improve the accessibility and reusability of fully digitized collections, some with insufficient data. To explore the possibilities of AI, *YLE* formed a multidisciplinary team involving the archive staff together with professionals from the editorial, operations, and multimedia departments. This team explored the possibilities of image recognition, scene segmentation, object and face recognition, and optical character recognition (OCR) on current affairs programs, as these were the most commonly used. This pilot project, carried out in 2016, did not provide a more time-efficient analysis and it also showed some limitations, such as the inability to recognize the identity of people appearing in images due to insufficient algorithm training.

That same year, they tested speech-to-text and content classification solutions for the radio archive. The results were good both in transcription and Named Entity Recognition, although they did not delve into speaker segmentation.

This project, far from offering real technological solutions, allowed the *Yle Archive* team to reflect on their actual needs, particularly in terms of defining which parts of the process can be automated and which should be addressed manually. *Yle*'s efforts are currently focused on the automatic generation of metadata for content production and web publication, automatic text generation from data and automatic subtitling. The ultimate goal is to have all the content produced by the broadcaster automatically analyzed with a particular emphasis on *Yle Areena*, the video-on-demand platform (**Viljanen**, 2022).

5.2.2. *ARD*

In 2017, the German consortium of public broadcasters, *ARD*, created a working group with the goal of identifying opportunities and creating use cases for the introduction of AI tools and methods in daily production. From the standpoint of the group, regionality and domain multiplicity are crucial elements in the implementation of AI solutions, also essential in the generation of appropriate metadata for new platforms, different users and content personalization or recommendations (**Wenger-Glemser**, 2019).

Among the broadcasters that are part of *ARD*, the Bavarian television (*Bayerischer Rundfunk* or *BR*) stands out for its use of AI. *BR* has developed a facial recognition model trained with its own data and use cases (**Schreiber**, 2022). In the initial phase of the project, the dataset to train the algorithm focused on scenes showing faces and identifying written signs extracted from two news programs, 30 and 15 minutes long. The result of this process was a demanding data model that required very high accuracy rates to incorporate the extracted metadata into the archive. By February 2022, *BR* had analyzed 55,000 images from the archive and detected 3,000 different classes. Human quality control was considered an essential element to detect false positives and the images responsible for them.

Along the same lines, *BR* has trained a model to identify historical buildings and relevant political and financial centers at a regional level (**Förster**, 2023). The data model, generated automatically from open-source tools, used 255 subtitled programs on which entity recognition and disambiguation processes were applied. This is a proof of concept that aims to be optimized and integrated with other solutions currently in use.

5.2.3. *Radio Télévision Suisse* (*RTS*)

Since 2018, *Radio Télévision Suisse* (*RTS*) has systematically approached the cataloging of their archive by applying AI techniques and developing an interface based on open-source technologies to transcribe audio to text, perform facial

recognition, and automatically classify images during the ingestion process into the archive (**Rezzonico**, 2020). Automatic classification has been successfully applied to uncategorized sports collections, allowing them to at least detect the topic of each recording. Facial recognition has been developed using a database of 5,000 Swiss public figures taking into account factors such as the duration of the shot and high reliability levels in order to avoid false positives, that is, the identification of one person as a different one because of their common facial traits (**Bouchet**; **Ducret**, 2019). Consequently, only faces identified with a reliability level above 85% are integrated into the archive. One of the main features of this tool is that it allows not only for the extraction of metadata but also for image retrieval or visual searching based on faces, scenes, monuments, and buildings. This functionality is particularly relevant for television archives when it comes to locating scenes that have been embargoed due to copyright issues or court orders. The *RTS.ai* interface has enabled the automatic speech to text transcription for a total of 10,000 hours within two years. *RTS* is currently working on the development of speaker recognition tools, the integration of speech-to-text transcription with facial recognition, the assessment of photograph and video aesthetic features and action description (**Sonderegger**, 2023).

5.2.4. *BBC*

The *British Broadcasting Corporation* (*BBC*) is one of the first European public broadcasters that has pondered on the consequences of applying AI to content production. *BBC 4.1* emerged as a project within the Research and Development department of the British public broadcaster with a dual objective: to understand how AI can influence the future of audiovisual production and to collaborate with content creators in order to be prepared for the future. Between September 4th and 5th 2018, the schedule of *BBC Four*, a channel specializing in cultural content, was generated by an algorithm which had been trained to detect the most relevant content for the channel based on programs broadcast in the past (*BBC*, 2018a) and analyzing program descriptions and topics covered. Out of a total of 270,000 programs, the algorithm identified 150 that it deemed to be the most relevant for broadcasting, which were later used to manually decide on the final schedule. This experience also led to the project "Made by machine: When AI met the archive" (*BBC*, 2018b), a series in four episodes created using AI to demonstrate how machines think. Techniques such as object and scene recognition, natural language processing (NLP) for subtitle analysis and dynamism were applied using the preselected 150 most relevant programs as basis. The result was broadcast on the channel in 2018 in four micro-programs with Hanna Fry, a British mathematician and popular science communicator, as presenter and contributions from *BBC* archivists.

Another project involving the use of AI at the *BBC* is "Starfruit Tagger" (**Berger**; **Armstrong**, 2022; *Starfruit Tagger*, 2023). The *BBC* has very few original recordings of the news bulletins aired during the first 50 years of its radio station's existence. However, it does have the scripts of these bulletins from 1937 to 1955, which have been digitized and processed using optical character recognition (OCR). A pre-trained automatic tagging system previously trained on sports and news content that had been manually indexed by the *BBC*'s editorial team has been used to extract names of people, places, organizations and events.

5.2.5. *TV2*

In 2021, the Norwegian commercial channel *TV2* began conducting tests on automatic speech-to-text transcription and subtitling for current affairs programs and raw material from news and programs (**Tverberg**, 2021). This project involved the collaboration between journalists and operations staff, who subjectively evaluated the quality of the generated transcription using services such as *Speechmatics* (*Speechmatics*, 2023), and *Azure* (*Microsoft*, 2023), among others, whose performance was also objectively measured by calculating their Word Error Rate (WER).

When asked about the quality of automatically generated subtitles as opposed to manually created ones, 9.1% of users considered them very good, 10.3% found them good, 21% believed the quality was sufficient, 39.4% thought they were of poor quality and 21.2% rated them as very poor quality. From the developers' perspective, the main challenges for these types of services are dialects, entity recognition, speaker segmentation, and WER, which, although decreasing over time, can still be significant depending on the context in which the error occurs.

TV2 has also conducted proof-of-concept tests with *CLIP* (2023), the neural network developed by *OpenAI*, on 5,000 elements from their archive with limited descriptive metadata (**Steskal**, 2023). The results obtained for person and object recognition were good despite the lack of specific training. However, they noticed that the keyframes returned by the system were not always representative of the entire video. Despite the good performance, they found that the system would not be able to handle complex information searches.

5.2.6. *VRT*

The Flemish public radio and television broadcaster, *VRT*, has invested in the development of its own AI models for the automatic scene segmentation of the content published on its website and the enrichment of archive metadata (**Daniels**; **Degryse**, 2021). The application of these solutions aims to improve efficiency, increase reusability, enable recommendations, and avoid repetitive tasks for the editing, multimedia, and archival teams. The project has been carried out in two phases: an initial preprocessing and training phase and an application phase.

During the first phase, the most relevant sources of information to train the segmentation algorithm were identified, which included information from broadcasts, subtitles, schedules, facial recognition and RGB (Red, Green and Blue) values. All the data gathered were used to create a vector representation and define scene transitions by measuring RGB

values. Finally, pairs of similar and dissimilar scenes were established. Once the algorithm was trained, in the application phase, program information was extracted and the model was applied. The results were validated by humans (**Daniels**; **Degryse**, 2021).

VRT also utilizes AI in production to enrich metadata through optical character recognition (OCR) applied to the lower third of on-screen images, where they manage to detect the speakers on screen. The information obtained through OCR is processed using natural language processing (NLP) to extract entities and keywords on which filters are applied to discard irrelevant information (**Daniels**, 2023).

5.2.7. *SVT*

Sveriges Television, the Swedish public television, known as *SVT*, has conducted several proof-of-concept tests in its production archive, *Mark*, which is also used as a platform to foster innovation within the company (**Åstrand**; **Ståhl**, 2023). These tests address the absence of metadata in a significant percentage of content and utilize AI techniques to make it accessible. Specifically, they have conducted image search tests using *CLIP* (facial recognition for local politicians), OCR to identify program creators and technical teams in credits, as well as automatic speech to text transcription and entity recognition to identify key individuals and topics discussed. The medium-term goal is to identify real use cases and bring them into production, incorporating only high-reliability metadata.

5.2.8. *Asharq News*

Asharq News (**Battrick**, 2022; **Battrick**; **Petitpont**, 2022), the multi-platform Arabic news network founded in November 2020, is presumably one of the few examples in the world where the use of AI was included as part of the network's initial development plan. The project, developed over 18 months, involved integrating their *AVID Media Central* production system with *Newsbridge* technology, aiming to generate metadata for 1,600 hours of monthly broadcast and original content both in English and Arabic. The proof-of-concept phase, prior to implementation, involved the participation of various user groups who not only contributed to the training of the facial recognition and automatic transcription and translation models, but also identified different use cases and defined the required accuracy rates for integration. The complexity of managing consistent metadata in English and Arabic, the challenge of training an automatic speech transcription model for a language with numerous dialects as Arabic, the network's specialization in political and military topics and the need to develop facial recognition models for Arabian prominent people earned *Asharq* international recognition with the *FIAT/IFTA Media Management Award* in 2022 (*FIAT/IFTA*, 2022).

5.2.9. *IRIB*

In the same context, it is worth noting the centralized data model and intensive use of AI by the Iranian television network, *IRIB*. The difficulty of finding data to train their own models has led this broadcaster to establish a data factory to develop projects for image and text labeling, verification or automatic pre-processing and data validation. These processes are then applied in both the production systems and the archive (**Ghanbari**, 2022).

This factory allows *IRIB* to develop its own tools to power various AI-based services for their archive, television channels and website, as well as for the administration and finance departments, which also leverage data generated by other areas. These tools are applied for text summarization and entity detection, video annotation and image summarization, speech-to-text transcription, text-to-speech conversion, gender detection, and speaker identification.

In Europe, *France Télévisions* and the *Italian Public Broadcasting Corporation*, *RAI*, have also embraced similar approaches.

5.2.10. *France Television*

DAIA is the data governance department at *France Télévision*, whose aim is to ensure the interoperability and availability of data for the various departments within the company. To achieve this, it has a knowledge interface that analyzes the data, shares them across different datasets and translates them when they lack coherence, in other words, it converts them into interoperable data. The foundation of this system is a common ontology that enables these datasets to mutually understand one another. Additionally, when the existing data is insufficient, open-source AI is applied to generate data based on the credits or the content itself (**Parmentier**, 2021).

5.2.11. *Radiotelevisione Italiana* (*RAI*)

In the case of *RAI*, the generation of datasets for machine learning is seen as a key element for the integration of AI into the workflows of the Italian public broadcasting company (**Messina**, 2021). In this regard, the metadata generated by the archive and the broadcasting and production areas, to name just a few, undergo a process of extraction, filtering, and adaptation to render them useful for model training. The *RAI Media Cognitive Service Platform* is the tool that enables data ingestion, content annotation using cloud solutions or proprietary models, data validation and enrichment, and the creation of refined collections suitable for the development of models that will be put into production. This means that *RAI* is independent from third parties in its AI-based development and is capable of generating and applying its own data models specifically adapted to its needs (**Messina**; **Montagnuolo**, 2023).

5.2.12. *Radio Televisión Española* (*RTVE*)

In Spain, *RTVE* and *Atresmedia* have been pioneers in the implementation of AI in their archives. In the case of *RTVE*, the first approaches were carried out through the creation of the *RTVE University of Zaragoza Chair* (*Cátedra RTVE Universidad de Zaragoza*, 2017). As part of the activities of this chair, the *RTVE Database* was first published in 2018. It consists of annotated datasets from TV programs and serves as the basis for the *RTVE Albayzin Challenges* (**Lleida-Solano** *et al.*, 2022). These challenges, which bring together national and international research groups, have allowed for the testing of state-of-the-art systems in automatic transcription and multimodal recognition with use cases prepared by the *RTVE* archive.

In the field of production, the tender for automatic metadata generation in the *RTVE* archive (*RTVE*, 2021) was awarded in 2021. This cloud-based service, through a technological integrator, enables automatic metadata generation for audio and video content spanning 11,000 hours of *RTVE* content. This service will be replaced in October 2023 by a new one that includes new functionalities such as automatic analysis and translation from Catalan, and where image analysis becomes more relevant (*RTVE*, 2023).

Furthermore, *RTVE* is working on a similar project for the *Radio Nacional de España archive* (*RNE*), which will enable transcription, automatic classification, and entity extraction for 190 hours of *Radio 1* and *Radio 5.*

5.2.13. *Atresmedia*

In 2019, *Atresmedia* launched a supervised automatic cataloging project. Through this on-premise service, the Atresmedia archive obtains transcription for 40 hours of daily content, including raw material, news items and fully subtitled programs. The generated metadata is integrated into the MAM (media asset management) system, where it is corrected and complemented with human cataloging. This project, which aims to transform the professional profile of documentalists from processors to content generators, was awarded, among others, with the *Excellence in Media Management Award* from *FIAT/IFTA* in 2021 (**López-de-Quintana**, 2021; **López-de-Quintana**; **León-Carpio**, 2021).

Other regional television broadcasters in Spain, such as *Aragón TV* (*Aragón Noticias*, 2021) and *Televisió de Catalunya CCM*, have carried out pilots of automated cataloging, with the latter incorporating automatic transcription in Catalan into its MAM system.

5.3. Press agencies

5.3.1. *Associated Press*

In the world of media, news agencies have not remained on the sidelines of technological progress. The *Associated Press* (*AP*), after a pilot and an 8-month development period, has integrated multimodal AI in collaboration with the Belgian company *Limecraft* in order to analyze both live feeds and recorded material which amount to approximately 700 video clips per month (**Coppejans**, 2021).

Multimodal technology integrates computer vision for scene identification and segmentation, facial and identity recognition, and role and attitude recognition on the one hand, with speech technologies that enable language detection and automatic transcription on the other. The results are integrated into an interface and displayed with the appropriate level of detail for *AP* (**Verwaest**, 2022).

5.3.2. *Reuters*

In 2020, *Reuters* announced the application of AI techniques to one million clips from its archive spanning from 1986 to the present (*Reuters Staff*, 2020). This project, funded by the *Google DNI Fund*, has enabled automatic speech-to-text transcription, translation into 11 different languages and recognition of internationally prominent people. It has also allowed *Reuters* to gain a better understanding of its archive, the analysis policies applied in the past and their effect on their content accessibility (*Reuters*, 2023). In this way, they have been able to improve their internal technological capacity and determine which types of content can be automatically analyzed with a high level of reliability and which cannot.

6. Joining forces: innovation projects

As we have seen, many radio and television companies have performed pilots with internal resources or implemented various AI-based solutions. However, in the last five years, we have also witnessed a significant rise in innovation projects where companies and media organizations join forces to advance the application of AI in their workflows.

6.1. *VIVA*

The *VIVA* project has led to the development of a tool to implement video retrieval methods based on deep learning models (**Mühling** *et al.*, 2022). Researchers from the *TIB - Leibniz Information Centre for Science and Technology* and the *University of Marburg* (Germany) have participated in the project together with *ARD* professionals. The objective is to enable concept-based or personality-based video retrieval in media archives and continuously update the deep learning model as new emerging personalities or needs arise. The tool has been tested on four use cases within the context of a collection of historical videos from the *German Broadcasting Archive*, which consists of approximately 34,000 hours of television recordings from the former German Democratic Republic.

6.2. *Europeana Subtitled*

Another relevant project in this field is *Europeana Subtitled*. This initiative has brought together a consortium of 7 European public television broadcasters, the *Fondazione Bruno Kessler* (*FBK*) and *Translated* (Italy) with the aim of developing automatic speech-to-text transcription, translation, and automatic subtitling models in order to improve the accessibility of audiovisual content in collections such as *Europeana* (**Lewis**; **Jarret**, 2023). This project has made it possible to upload 8,000 English-subtitled videos focusing on the topic "Broadcasting Europe" onto the European digital library, as a means to showcase the social changes that have taken place in Europe since the 1930s.

6.3. *AI4Media*

The previously mentioned *AI4Media* project aims to develop innovative tools to address the current challenges in the media sector (*AI4Media*, n.d.). To achieve this, it has defined seven industrial use cases that range from social media and misinformation to supporting newsrooms through automatic news creation. One notable use case is the application of computer vision techniques and automatic metadata generation to archival material in order to support news coverage of unexpected events, where immediacy and quality do make a difference. The project involves 9 universities, 9 research centers and 12 companies in the sector, including heritage preservation organizations, radio and television companies and vendors.

6.4. *Tailored Media*

Tailored Media is an innovation project led by *Joanneum Research* in collaboration with *ORF* and *Austrian Mediatheque* (**Bailer**; **Bauer**; **Rottermanner**, 2021) whose objective is to automatically generate relevant metadata using computer vision and natural language processing techniques, integrate them into current workflows and develop user-oriented processes based on these metadata.

7. Results of the *FIAT/IFTA* survey on the use of AI in media archives

In April 2023, *FIAT/IFTA* conducted a survey on the use of AI in media archives (*FIAT/IFTA*, 2023). This survey aimed to provide an overview of the degree of implementation of AI. More specifically, it sought to understand the technologies and applications being used, the level of their implementation, how metadata is being integrated into archive management systems and the expected future evolution of content cataloging.

A total of 54 organizations participated in the survey, primarily television networks and national or regional audiovisual archives in Europe. Out of the 54 respondents, 61% (33 organizations) stated that they are currently applying AI in their archives, 9% (5 organizations) reported using it exclusively in the production archive and 28% (15 organizations) have plans to use AI in the future, while only 2% (1 organization) do not anticipate using it in the future.

Based on the responses obtained, it can be stated that organizations are using or planning to use a combination of different technologies with a slight preference for audio and speech technologies (37%, 33) in the first place, followed by computer vision (34%, 31) and finally natural language processing (29%, 26). If we break down the data by technologies in production or in the planning phase, the trend remains the same for those organizations planning to integrate AI in the future, while there is a higher use of computer vision (27) as opposed to audio and speech technologies (25) and natural language processing (20) among the organizations who are already applying AI.

Graph 1 displays the degree of use of specific applications related to speech and audio technologies. Graph 2 showcases the specific applications related to natural language processing. Graph 3 exhibits the percentage of usage for applications related to computer vision.
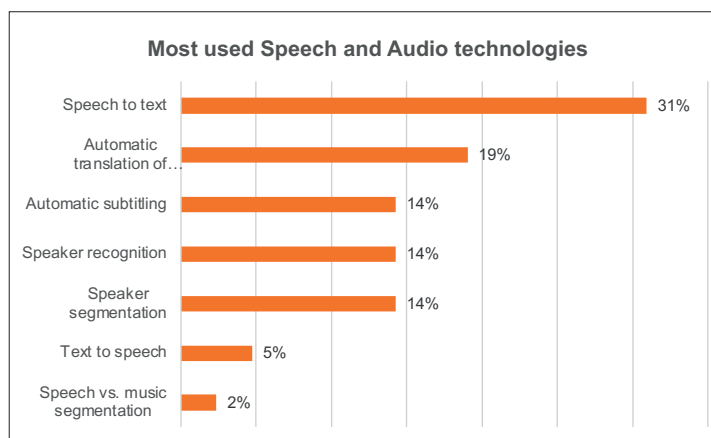
Graph 1. Use of speech and audio technology applications (%).
Source: *FIAT/IFTA* (2023).
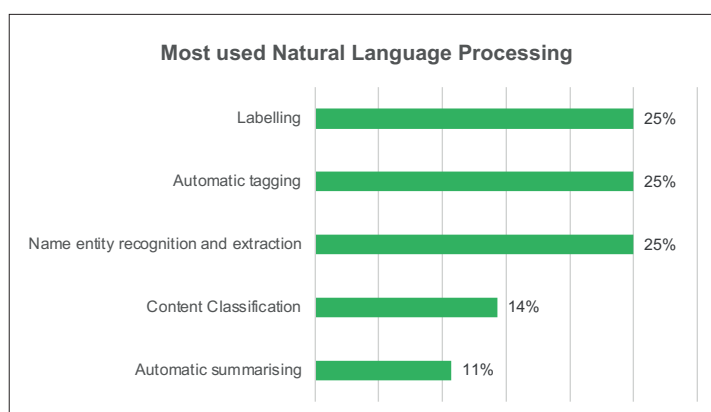
Graph 2. Use of natural language processing applications (%).
Source: *FIAT/IFTA* (2023).

In a significant number of cases, these technologies are fully integrated into the production process, while in a smaller percentage, they are used in upcoming innovation projects or proof of concept to understand the scope and limitations of the technology.

Another relevant factor highlighted in this survey is the origin of the technologies currently being implemented by media archives. 26% (13) use open-source solutions implemented by their own organization, another 26% (13) use third-party technologies (such as *Azure*, *Amazon*, *IBM*, etc.) implemented by their own organization, 26% (13) rely on third-party services that integrate proprietary technologies, and 23% (11) also rely on third parties that in turn integrate technologies from others.

Graph 3. Use of applications related to computer vision (%).
Source: *FIAT/IFTA* (2023).

Regarding how data is presented to the end user in archive management systems, 57% (24) do not indicate the source of the metadata, 12% (5) indicate it in pop-up windows, and 31% (13) indicate it on the same data display screen. Finally, the majority of organizations that responded to the survey believe that manual cataloging will decrease by 25% to 50% in the next 5 years.

8. Conclusions

Since 2007, media archive professionals have been contemplating the application of AI-related technologies and the future evolution of their work, particularly in activities such as content annotation. The initial AI projects have been driven by the innovation departments within companies and developed by multidisciplinary teams. In many cases, start-ups, research groups or specialized companies have been involved, such as *Yle*, *RTVE*, and *RSI*. Some organizations have opted for customized developments based on open-source solutions, especially from 2019 onwards.

The main objective of applying AI is to improve process efficiency, increase content reusability, enable website recommendations and avoid redundant tasks performed by different user groups that generate the same information about the same content at different stages of the production chain (*VRT*, *IRIB*, *RAI*, *France TV*). Speech and audio technologies can be considered widely implemented, although there are challenges to overcome, particularly those related to the lack of datasets to develop language models in low-resource languages. The development of models based on well-structured proprietary data coexists with the use of technologies provided by third parties, whether they are proprietary or commercial technologies adapted to the specific needs of broadcasters.

The application of computer vision techniques currently focuses on facial recognition in controlled collections, aiming to identify public figures of interest, particularly in regional contexts, where the solutions provided by major technology companies may not be sufficient. In these cases, models with high accuracy rates are employed to prevent false positives, while strict data protection policies are implemented to avoid using images of children or non-public individuals (*NISV*, *RSI*, *BR*).

The use of AI to automatically generate content from archival material is currently anecdotal or limited in scope.

Lastly, the application of AI in archives poses significant challenges in terms of scalability, integration with production and archival systems, and the restructuring of work for documentalists in both analysis and user support units. Additionally, the integration of metadata into archival management systems is a complex task that has a considerable impact on searches and selection results for both expert users and end users who increasingly rely on these tools without professional intermediaries.

In conclusion, the use of AI in media archives presents significant opportunities to enhance accessibility and new content productions based on archival material. However, it also raises important issues regarding its integration into workflows, data quality and reliability. As its application expands, conducting further research and fostering collaboration among professionals from various disciplines will be crucial to harness the full potential of AI while ensuring its responsible integration.

9. References

AI4Media (2023). *The AI4Media project*.
https://www.ai4media.eu

Aragón Noticias (2021). "Aragón TV participa en la prueba piloto de una herramienta de inteligencia artificial". *Aragón noticias*, 13 noviembre.
https://www.cartv.es/aragonnoticias/sociedad/aragon-tv-participa-en-la-prueba-piloto-de-una-herramienta-de-inteligencia-artificial-6130
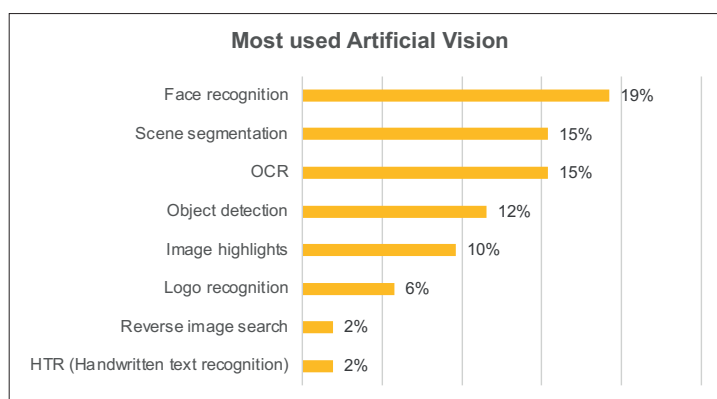
Åstrand, Mikaela; **Ståhl, Sally** (2023). "Finding without tagging: AI experiments for improved findability in the new media archive at SVT". In: *EBU Data technology seminar 2023*.
https://tech.ebu.ch/publications/ai-experiments-for-improved-findability-in-the-svt-media-archive

Bailer, Wermer; **Bauer, Christoph**; **Rottermanner, Gernot** (2021). "Analysing used needs for automatic metadata creation and advanced search in the Tailored Media Project". In: *EBU MDN workshop 2021*.
https://tech.ebu.ch/publications/analysing-user-needs-for-automatic-metadata-creation-and-advanced-search-in-the-tailored-media-project

Battrick, Kathey (2022). "Case study: AI indexing. Assessing and selecting AI services". In: *FIAT/IFTA world conference: Archive out of the box!*
https://fiatiftaworldconference2022.sched.com/event/149Cd

Battrick, Kathey; **Petitpont, Frederic** (2022). "How Asharq News is leveraging multimodal AI indexing with Newsbridge: Inside the back-end, operational & technical workflows of a multilingual news channel". In: *FIAT/IFTA world conference: Archive out of the box!*
https://fiatifta.org/world-conference-2022-recordings-available

BBC (2018a). *AI TV on BBC 4.1: BBC Four is giving an AI control of your TV for two nights*.
https://www.bbc.co.uk/programmes/p06jrfcc

BBC (2018b). *Made by machine: When AI met the archive*. BBC Four.
https://www.bbc.co.uk/programmes/b0bhwk3p

Berger, Jake; **Armstrong, Andy** (2022). "BBC radio news scripts 1937-1995: Using an automated tagger to enable journeys across time and space". In: *EBU MDN workshop 2022*.
https://tech.ebu.ch/publications/bbc-radio-news-scripts-1937--1995-using-an-automated-tagger-to-enable-journeys-across-time-and-space

Bouchet, Leonard; **Ducret, Sebastien** (2019). "A visual feature extraction pipeline and its applications for radio television Suisse". In: *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles, part IX*, pp. 51-59.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

Bruccoleri, Angelo; **Iacoviello, Roberto**; **Messina, Alberto**; **Metta, Sabino**; **Montagnuolo, Maurizio**; **Negro, Fulvio** (2022). *AI in vision: High quality video production & content automation*. RAI. Radiotelevisione Italiana. Centre for Research, Technological Innovation and Experimentation. AI4media.
https://www.ai4media.eu/whitepapers/ai-in-vision-high-quality-video-production-content-automation

Cátedra RTVE Universidad de Zaragoza (2017).
https://catedrartve.unizar.es

Coppejans, Charlotte (2021). *How artificial intelligence slashes editing time for Associated Press (AP)*.
https://www.limecraft.com/how-the-associated-press-is-using-ai-to-create-automated-shot-lists

Couteux, Anne; **Segura, Olivio** (2023). "News channel automatic segmentation". In: *FIAT/IFTA media management seminar 2023*.
https://fiatifta.org/seminar/media-management-seminar-2023

Daniels, Marijin (2023). "Metadata enrichment using lower third character recognition". In: *FIAT/IFTA media management seminar 2023*.
https://fiatifta.org/seminar/media-management-seminar-2023

Daniels, Marijn; **Degryse, Jasper** (2021). "How VRT automated the segmentation of programmes with AI". In: *EBU MDN workshop 2021*.
https://tech.ebu.ch/publications/how-vrt-automated-the-segmentation-of-programs-with-ai

FIAT/IFTA (2022). *Archive Achivement Awards 2022*.
https://fiatifta.org/awards-2022

FIAT/IFTA (2023). *AI Survey. The results are now available*.
https://fiatifta.org/ai-survey-results

FIAT/IFTA Media Management Commission (2013). *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles. Part VI. Metadata as the cornerstone of digital archiving*.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

FIAT/IFTA Media Management Commission (2017). *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles. Part VIII. Embracing automation – enhancing discoverability*.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

FIAT/IFTA Media Management Commission (2019). *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles. Part IX. Game changers? From automanation to curation - Futureproofing AV content.*
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

FIAT/IFTA Media Management Commission (2023). *MMC Seminar. Download the slides.*
https://fiatifta.org/mmc-seminar-2023-download-the-slides

Förster, Constantin (2023). "Generation of training data for landmark recognition in videos using named entity recognition". In: *EBU data technology seminar 2023.*
https://tech.ebu.ch/publications/landmarkner--generating-training-data-for-landmark-recognition

Ghanbari, Shirin (2022). "Strategy and challenges for developing AI applications". In: *EBU AIM community meeting 2022.*

Green, Eva-Lis; **Gupta, Jacqui** (2019). "20 year of MMC seminars: changing sceneries, changing roles 1998-2019". In: *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles, part IX*, pp. 7-17.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

INA (2023a). *International affairs.* Institut National de l'Audiovisuel.
https://www.ina.fr/institut-national-audiovisuel/international-affairs

INA (2023b). *Collection, preservation, and documentation of audiovisual heritage.* Institut National de l'Audiovisuel.
https://www.ina.fr/institut-national-audiovisuel/collection-preservation-and-documentation-of-audiovisual-heritage

Lewis, Michelle; **Jarret, Nicholas** (2023). *Explore how Europeana Subtitled increased access to audiovisual heritage.*
https://pro.europeana.eu/post/explore-how-europeana-subtitled-increased-access-to-audiovisual-heritage

Lleida-Solano, Eduardo; **Ortega-Giménez, Alfonso**; **Miguel, Antonio**; **Bazán-Gil, Virginia**; **Pérez-Cernuda, Carmen**; **De-Prada, Alberto** (2022). *RTVE 2018, 2020 and 2022 database description.*
http://catedrartve.unizar.es/reto2022/RTVE2022DB.pdf

López-de-Quintana, Eugenio (2021). *AI algorithms for media cataloguing in Atresmedia Group.*
https://fiatifta.org/awards-2021

López-de-Quintana, Eugenio; **León-Carpio, Antonio** (2021). "Artificial intelligence for a role change in television archives: The *Atresmedia–Etiqmedia* experience". *Journal of digital media management*, v. 10 n. 2, pp. 177-187.
http://etiqmedia.com/ficheros/JDMM_10_2_JDMM0006_e_lopez_de_quintana_and_carpio-1.pdf

Manders, Tim (2019). "Harder, better, faster, stronger: Adding face recognition in the mix at NISV". In: *FIAT/IFTA world conference: AV archives in the all-media world.*
https://fiatifta2019.sched.com/event/S0Sa/harder-better-faster-stronger-adding-face-recognitionin-the-mix-at-nisv

Manders, Tim (2022). "Face recognition at the *Netherlands Institute for Sound and Vision*. Is it really harder, better, faster, stronger". In: *FIAT/IFTA media management seminar.*
https://fiatifta.org/seminar/media-management-webinars-2022-2

Manders, Tim; **Wigham, Marie** (2021). "More metadata, lots of links, but what do you do with them? Practical examples of the added value of automatic metadata and linked data for archive users". In: *2021 FIAT/IFTA world conference. Advancing the digital dividend.*

Martin, Camille; **Segura, Olivio** (2021). "Using AI tools to segment and describe broadcast live stream. In: *EBU MDN workshop 2021.*
https://tech.ebu.ch/publications/using-ai-tools-to-segment-and-describe-broadcast-livestream

Messina, Alberto (2021). "Enabling AI with dataset engineering". In: *EBU MDN workshop 2021.*
https://tech.ebu.ch/publications/enabling-ai-with-dataset-engineering

Messina, Alberto; **Montagnuolo, Maurizio** (2023). "Integrating open knowledge-bases and AI-tools". In: *EBU data technology seminar 2023.*
https://tech.ebu.ch/publications/integrating-open-knowledge-bases-and-ai-tools

Microsoft (2023). *Azure cognitive services documentation.*
https://learn.microsoft.com/en-us/azure/cognitive-services

Mühling, Markus; **Korfhage, Nikolaus**; **Pustu-Iren, Kader**; **Bars, Joanna**; **Knapp, Mario**; **Bellafkir, Hicham**; **Vogelbacher, Markus**; **Schneider, Daniel**; **Hörth, Angelika**; **Ewerth, Ralph**; **Freisleben, Bernd** (2022). "VIVA: visual information retrieval in video archives". *International journal on digital libraries*, v. 23, n. 4, pp. 319-333.
https://doi.org/10.1007/s00799-022-00337-y

Netherlands Institute for Sound and Vision (2023). *Mission and vision.*
https://www.beeldengeluid.nl/en/about/mission-and-vision

Open AI (2022). *Introducing ChatGPT.*
https://openai.com/blog/chatgpt

Open AI (2023). *Dalle-E 2.*
https://openai.com/dall-e-2

Parmentier, Matthieu (2021). "The challenge of data management and governance". In: *EBU MDN workshop 2021*.
https://tech.ebu.ch/publications/governance-of-data-at-ftv-for-better-usage-and-valorisation-of-the-programs

Petit, Thomas (2022). "Facial recognition: Trombinos". En: *FIAT/IFTA media management seminar.*
https://www.youtube.com/watch?v=qM10EXb6988

Reuters (2023). *Automated transcription, translation and identification.*
https://liaison.reuters.com/help/reutersconnect#automated-transcription-translation-and-identification

Reuters Staff (2020). *Reuters applies AI technology to 100 years of archive video to enable faster discovery, supported by Google DNI.*
https://www.reuters.com/article/rpb-lavita-video-archive/reuters-applies-ai-technology-to-100years-of-archive-video-to-enable-faster-discovery-supported-by-google-dni-idUSKCN2591VO

Rezzonico, Pietro (2020). "AI-based tools to help archive workflows". In: *EBU tech roundtable on archives 2020*.
https://tech.ebu.ch/publications/presentations/roundtable_2020_archive_workflows/RTS-archive_tools

Roche-Dioré, Axel (2023). "data.ina.fr a portal to promote media analytics". In: *EBU data technology seminar 2023.*
https://tech.ebu.ch/publications/datainafr--a-portal-to-promote-media-analytics

RTVE (2021). *Metadatado automático de contenidos del Fondo Documental de RTVE.*
https://licitaciones.rtve.es/licitacion/licitaciones/detalle?id=1208797

RTVE (2023). *Metadatado automático de contenidos del Fondo Documental RTVE*
https://licitaciones.rtve.es/licitacion/licitaciones/detalle?id=2200005

Sánchez-García, Pilar; **Merayo-Álvarez, Noemí**; **Calvo-Barbero, Carla**; **Diez-Gracia, Alba** (2023). "Spanish technological development of artificial intelligence applied to journalism: companies and tools for documentation, production and distribution of information". *Profesional de la información*, v. 32, n. 2.
https://doi.org/10.3145/epi.2023.mar.08

Schreiber, Jonas (2022). "AI going local - conception and training of specialized Bavarian' AI models at Bayerischer Rundfunk". In: *EBU AI-AME 2022*.

Selkälä, Elina (2017). "Heading for AI". In: *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles, part VIII. Embracing automation - Enhancing discoverability*, pp. 169-189.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars

Sonderegger, Janique (2023). "Speaker recognition service: Roles and expectations". In: *FIAT/IFTA media management seminar.*
https://fiatifta.org/seminar/media-management-seminar-2023

Speechmatics (2023). *The most accurate speech to text API.*
https://www.speechmatics.com

Startfruit Tagger (2023). *BBC Startfruit tagger.*
http://starfruit.virt.ch.bbc.co.uk/#

Steskal, Lubos (2023). "Using free-text queries to perform visual search in video archives". In: *EBU data technology seminar 2023.*
https://tech.ebu.ch/publications/using-free-text-queries-to-perform-visual-search-in-video-archive

Tverberg, Are (2021). "TV 2 Norway's AI hub and QA Monitor: A gateway to ASR services for a low-resource language". In: *EBU MDN workshop 2021*.
https://tech.ebu.ch/publications/tv-2-ai-hub-and-qa-monitor--a-gateway-to-asr-services-for-a-low-resource-language

Verwaest, Maarten (2022). "Using AI for automatic shot listing in news at *Associated Press*". In: *EBU MDN workshop 2022.*
https://tech.ebu.ch/publications/using-ai-for-automatic-shot-listing-in-news---the-associated-press

Viljanen, Kim (2022). "Content tagging in *Yle Areena*: Improving quality and process". In: *EBU MDN workshop, 2022.*
https://tech.ebu.ch/publications/content-tagging-in-yle-areena---improving-quality-and-process

Wenger-Glemser, Gabriele (2019). "Shaping the future Artificial Intelligence in AV archives". In: *Proceedings of the FIAT/IFTA media management seminars. Changing sceneries, changing roles, part IX*. pp. 61-78.
https://fiatifta.org/library/proceedings-of-the-fiat-ifta-media-management-seminars