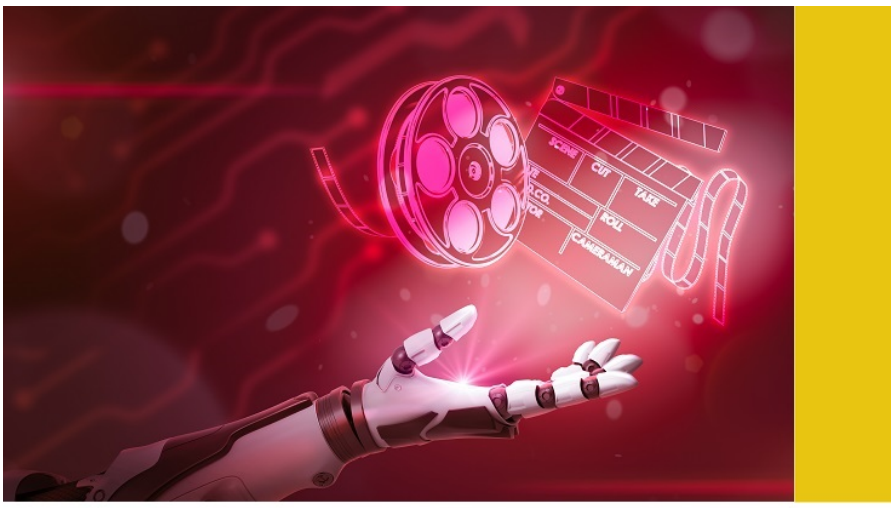
IN-DEPTH ANALYSIS
Requested by the CULT Committee

European Parliament

Research for CULT Committee - The use of Artificial Intelligence in the Audiovisual Sector

Concomitant expertise for INI report



Culture and Education

Policy Department for Structural and Cohesion Policies
Directorate-General for Internal Policies
PE 629.221 - May 2020

EN

RESEARCH FOR CULT COMMITTEE

The use of Artificial Intelligence in the Audiovisual Sector

Concomitant expertise for INI report

Abstract

This paper contributes towards the identification of current activities, important priority topics and crucial needs in the area of Artificial Intelligence (AI) within the European audiovisual sector. This document was provided by the Policy Department for Structural and Cohesion Policies on the request of the CULT Committee in order to support the drafting of the own-initiative report (INI) "The use of artificial intelligence in the education, culture and audio-visual sectors".

This document was requested by the European Parliament's Committee on Culture and Education.

AUTHOR

Georg REHM

Research administrator: Katarzyna Anna ISKRA
Project and publication assistance: Lyna PART
Policy Department for Structural and Cohesion Policies, European Parliament

LINGUISTIC VERSIONS

Original: EN

ABOUT THE PUBLISHER

To contact the Policy Department or to subscribe to updates on our work for the CULT Committee please write to: Poldep-cohesion@ep.europa.eu

Manuscript completed in May 2020

This document is available on the internet in summary with option to download the full text at: https://bit.ly/2EBMd1a

This document is available on the internet at:
http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_IDA(2020)629221

Further information on research for CULT by the Policy Department is available at:
https://research4committees.blog/cult/
Follow us on Twitter: @PolicyCULT

Please use the following reference to cite this study:
Rehm Georg, 2020, Research for CULT Committee – The use of Artificial Intelligence in the Audiovisual Sector, European Parliament, Policy Department for Structural and Cohesion Policies, Brussels
Please use the following reference for in-text citations:
Rehm (2020)

DISCLAIMER

KEY FINDINGS

This briefing paper takes a look at the use of AI technologies in the wider audiovisual sector. A survey with ten questions concerning the most important aspects was circulated to 85 contacts at 73 organisations. A total of 22 responses were received. The main findings are as follows:

1. Almost all respondents report broad use of AI technologies, especially for automated indexing, improved content accessibility as well as localisation. AI is used for processing audio or video, language or text data or for knowledge management purposes.
2. Among the technologies used are ASR, TTS, NLP, NER, MT, summarisation, search and recommender engines, content classification, subtitling, vision and metadata extraction (see Appendix 2: Glossary – Terms and Abbreviations).
3. AI technologies foreseen for future use are more experimental and include the automated detection of illegal content and deep fakes as well as flexible curation technologies.
4. There is a big demand for large amounts of training data including labelled, structured and unstructured data, domain-specific training data, acoustic data and data for illegal content.
5. There is also a need for more language technologies for all European languages, including ASR, TTS, MT, content curation services and metadata extraction as well as Linked Data.
6. In terms of policies, it is suggested to focus upon an ethical framework regarding the use and misuse of AI that protects human values and fosters cultural and linguistic diversity. It should also protect against the misuse of AI for false news and misinformation.
7. Regarding opportunities, many respondents suggest concentrating on the AI-based production of high quality content. In addition, AI allows unlimited localisation and makes it possible for a fragmented and culturally diverse ecosystem to survive in a world dominated by capital-intensive ventures based in the US.
8. The consumption of intentionally created false or manipulative content is seen as an imminent danger. It is stressed that, as video is quickly becoming our main means of communication, there is a threat that relates to the use of AI for misinformation and manipulation, which could have an impact on the foundations of our democratic society.
9. The awareness of the European AI tool market varies. Some perceive the market to be non-existent, others perceive it to be highly fragmented. Due to the dominance of non-European technology enterprises, European companies should be supported more.
10. Collaboration at the European level is seen as essential because individual players have limitations and difficulties in using AI technologies. Europe's multilingualism is seen as crucial: to guarantee inclusiveness and accessibility, tools need to be made available, especially for under-resourced languages.

The briefing paper concludes with the following four recommendations:

1. Support long-term European platform initiatives.
2. Reinforce the European AI and Language Technology providers landscape.
3. Support the European Digital Public Space Vision (European Media Platform).
4. Conduct an in-depth study of the sector.

Introduction: AI in the Audiovisual sector

This briefing paper is a contribution towards the identification of current activities, important priority topics and crucial needs in the area of Artificial Intelligence (AI) within the European audiovisual sector. The paper highlights the main topics and makes a number of long-term policy recommendations. Specifically, the paper concentrates on the impact and relationship of AI-based technologies with regard to the creation, production and consumption of audiovisual and multimedia content. It contributes to the identification of AI-driven use cases in the sector and characterises important aspects of the market of tool providers. It emphasises important components of the AI-driven value chain for audiovisual content and touches upon issues raised by the increased use of AI.

AI in the Audiovisual Sector: Potential scope

While AI itself is a broad and diverse field of research whose origins date back to the 1950s, the term is currently used almost exclusively for technologies developed with different forms of Machine Learning (ML), especially neural network-based approaches, often referred to as Deep Learning (DL). It is this powerful family of neural technologies especially that has, in recent years, contributed to breakthroughs in many applications, from robotics, to image and audio processing and to automated language analysis. As a consequence, the intersection between AI-based technologies on the one side and the audiovisual sector (including film, television, video games and multimedia) on the other, is extensive. Among many other fields of application, this could refer to: AI for camera technologies (picture stabilisation etc.); software for manipulating, editing, processing, enriching textual content, images, movies or audio files; Augmented Reality (AR), i.e., adding virtual images to real ones, anchoring artificial images to objects in a movie; deep fake detection: analysing movies, audio or text files to determine if they have been intentionally manipulated; AI-based television programme planning, i.e., user model-based dynamic programming; recommender systems for media repositories; (semi-)automated localisation and translation of content (video, audio, text) into other languages; autonomous non-player characters in video games; automated transcription and subtitling of video content; enhanced movie production using AI technologies. These are only a few selected examples of where AI and media meet, many more could easily be added.

AI in the Audiovisual sector: Scope of this paper

This briefing paper is based on a rather broad interpretation of "audiovisual sector", which includes not only the areas of audio and video but also the area of textual content, which is an inherent additional modality in nearly all audiovisual use cases, especially with regard to the (ubiquitous) multimedia applications in the World Wide Web and mobile apps, which combine text, video and audio seamlessly. As a consequence, and also because these different modalities (audio, video, text) are converging more and more, in addition to the areas and domains of film, television and games, we also include online publishing, content curation (including localisation) and multimedia. With regard to AI-based technologies, the key intersection of all of these different areas and fields is *language* itself, which is an inherent part of video content, audio content and textual content and for which various AI-driven technologies exist, which are usually referred to as Language Technologies (LT) or Language-centric AI. This area is divided between the analysis or generation of (a) written, i.e., textual language (Natural Language Processing, NLP) or (b) spoken language (Speech Technologies).

Survey

This briefing paper is primarily informed by a specifically prepared survey (cf. appendix). The survey contains ten questions (see below), tailored to the specifications set out in the terms of reference. Its main goal was exploratory, i.e. to map the main topics and trends in terms of the use of AI in the audiovisual sector. Accordingly, and to enable the respondents to answer in a flexible way, the survey consisted almost exclusively of free text fields (in contrast to multiple choice questions) with length limits. The survey was circulated to 85 contacts at 73 organisations, all of which are stakeholders of the wider audiovisual sector. The contacts were carefully selected from the author's own list of contacts (approx. two thirds of the contacts) and representatives from important stakeholders identified through desk research (approx. one third of the contacts). The survey was circulated via email on 6 March 2020 with a response deadline of 11 March 2020 (three workdays). The following stakeholders were included: public and private broadcasters, film business, video game developers, publishers and publishing houses, telecom providers, technology service providers, umbrella associations, research and academia.

Responses: General Overview

Despite the very short deadline and the rapidly developing Covid-19 pandemic, a total of 22 responses were received. This unexpectedly high response rate demonstrates the commitment of the respondents. Responses were received from Germany (11), France (3), Belgium (2), Latvia (2), Greece, Spain, The Netherlands and the UK (1 each). Representatives of the following stakeholders responded: technology or service provider (8), public broadcaster (3), public broadcast archive (2), research (2), consultancy firm, cultural organisation (philanthropy), film business, news agency, private investment network, public broadcast research institution, publishing association (1 each). The respondents have the following job titles/profiles: Director, Deputy Director, Head of Department, Programme Manager (7), CEO (5), Head of Research (3), AI/ML Lead (3), Head of Innovation (2), Other (2).

List of Questions

Q1. **Current use of AI:** Do you currently use AI-based technologies in your organisation?

Yes (in production use), Yes (experimental use), No, Prefer not to answer

If yes: In which areas do you apply AI-based technologies? If you use AI in several different areas, please feel free to focus on those areas that you perceive to be the most important ones.

Q2. **Future use of AI:** In addition to your answer in Q1, do you have plans to use AI-based technologies in your organisation in the next 24 months?

Yes (in production use), Yes (experimental use), No, Prefer not to answer

If yes: Please list the planned areas of application of AI-based technologies. If you plan to use AI in several areas, please feel free to focus on those areas that you perceive to be the most important ones.

Q3. **Needs, demands and requirements:** What are your organisation's concrete needs, demands and requirements in terms of AI-based technologies (and in which areas)?

Q4. **EU policies:** Do you see a need for establishing EU policies (including regulation and recommendations) for the use of AI-based technologies in the audiovisual sector and why?

Q5. **Opportunities of AI in the content life-cycle:** AI-based technologies can be used for the creation, production or consumption of audiovisual content. In which of these three phases of the content life-cycle do you see the biggest opportunities for AI-based technologies and why?

Q6. **Dangers of AI in the content life-cycle:** AI-based technologies can be used for the creation, production or consumption of audiovisual content. In which of these three phases of the content life-cycle do you see the biggest danger of AI-based technologies and why?

Q7. **European AI tool market:** Please provide a short description of how you assess the market of European AI tool providers for the audiovisual sector (for example, in terms of market access, market size, market diversity, pricing, competitiveness, market dominance etc.).

Q8. **Opportunities of AI (in general):** Independent of your organisation, in which areas do you see the biggest potential or set of opportunities (incl. technical, social, ethical, economic, legal aspects) for AI-based technologies in the audiovisual sector?

Q9. **Dangers of AI (in general):** In which areas do you see the biggest dangers or threats (incl. technical, social, ethical, economic, legal) of AI-based technologies in the audiovisual sector?

Q10. **European collaboration:** In which broader areas regarding the use of AI-based technologies in the audiovisual sector do you see a need for increased European collaboration or coordination?

Survey: Detailed Analysis of the Responses

In the following we examine the reponses to the ten questions, clustered by stakeholder group.

Q1: Current use of AI – Do you currently use AI-based technologies in your organisation?

All eight **technology and service providers** currently offer (or use) AI-based services. AI is used in three broad fields: processing (1) audio or video data, (2) language or text data or (3) for information and knowledge management purposes. In the audio field, AI is used for automatic speech recognition (ASR) and text to speech systems (TTS), and in the video field AI is used for the automation of post-production (auto rough cuts), and also for the processing of raw footage. Furthermore, AI can be used for speaker classification, emotion recognition and acoustic scene detection (audio analytics). For language and text processing, AI is used for various functionalities relating to the processing of textual content including named entity recognition (NER), machine translation (MT), optical character recognition (OCR), automatic concept and meaning extraction, internal search and recommender engines, content classification, automatic subtitling and localisation. This area of Language Technologies overlaps with Information and Knowledge Technologies, with knowledge graphs and knowledge bases playing a particularly important role in the media area, especially for the management of metadata. These technology providers offer services that are often used by the next stakeholder group, **public broadcasters**. Two of the three public broadcasters use AI in production and also in an experimental setting. Both mention almost exactly the same services using technologies that are rather mature and that have been in use in the area for several years: transcription of audio (ASR), translation of text content (MT), speech generation (TTS) and automated text summarisation, especially for automated subtitling of news content. Together with MT (experimental), summarisation is also applied to different languages and not only for the production of content but also for the multilingual monitoring of social networks and social media. Two of the broadcasters use computer vision technologies, i.e., object, face or scene detection to automatically extract additional metadata from video footage for automated indexing purposes. The third public broadcaster uses AI only in production. In addition to the above, this broadcaster uses AI to enhance video encoding and content delivery. The two **public broadcaster archives** are experimenting with AI-based technologies for nearly identical

purposes, i.e. smart search, face recognition and image classification (for metadata extraction and indexing), for segmenting broadcast news streams into individual parts and for video quality enhancement and upscaling. Knowledge graphs are used for the management of semantic metadata (for a video clip this can include, among others, *author*, *editor*, *date/time* recorded, *genre*, *language*, *keywords*, *summary*, *people* shown in the video including timecodes, *objects* shown in the video including timecodes etc.), especially in the Linked Data paradigm, which enables the linking of metadata to the corresponding data they describe using the World Wide Web stack of technologies. One **public broadcaster research organisation** participated in the survey. They experiment with ASR, the automated analysis and annotation of video data and NER for metadata generation. One of the two **research** organisations develop AI systems for production and experimental use. They provide an AI framework for indexing audio or video data including speech, video and text mining, which is used in production in a German public broadcaster. While most of the other stakeholders do not use AI, one larger organisation, a news agency, uses MT to translate image search queries from other languages into English. They also use face recognition on images taken at big events to facilitate and speed up the process of tagging pictures with person names (indexing and metadata generation).

Q2: Future use of AI – In addition to your answer in Question 1, do you have plans to use AI-based technologies in your organisation in the next 24 months?

Seven of the eight **technology and service providers** plan to use AI in production in the near future. Generally speaking, all of these planned future uses of AI technologies are slightly more advanced than the ones mentioned in Q1. Among others, the technology providers plan to provide services for the automatic detection of illegal content (weapons, drugs, illegal organisation logos etc.) in text, images, video and audio. Also mentioned were various types of content curation technologies, for example, exposing new pieces of factual information in documents to knowledge workers or to use AI for better support workflows, for example, through chatbots and automated question answering. In stark contrast, the **public broadcasters** mostly focus on improving and enhancing the technologies they already mentioned in Q1. They all plan to use AI in production in the next 24 months including tools for the automated subtitling of on-demand videos and (semi-)live news and closed captioning for improving accessibility. The two **public broadcaster archives** also want to improve their existing AI-based tools, taking them from an experimental phase to production. The **broadcaster research organisation** will be experimenting with NER and topic detection on top of automatically created transcripts of audio or video data, especially for subtitle generation. Of note is the news agency mentioned above, which is attempting to use AI for the detection of deep fakes and synthetic media and also for predicting the location of photos.

Q3: Needs, demands and requirements – What are your organisation's concrete needs, demands and requirements in terms of AI-based technologies (and in which areas)?

The **technology and service providers** and also the **research** organisation have a clear demand for different types of training data for AI-based machine learning algorithms including acoustic data, labeled structured and unstructured data to enrich internal and external knowledge graphs, domain-specific training data and training data for illegal content. Better language technologies with a broader coverage (including official or unofficial European languages with smaller numbers of speakers) are also needed by several service providers. The same holds for knowledge graphs and terminologies for certain domains, i.e. structured semantic vocabularies that can be used for metadata enrichment. Furthermore, better technologies (higher quality, higher precision) are demanded including ASR for all EU languages (plus Russian, Arabic, Chinese) and pre-trained models for computer vision techniques. The **public broadcasters** mention some of these aspects as well, especially the best possible language technology engines for all European languages including ASR, NER and related tools. In addition, they focus on ethical guidelines and regulations, climate friendly services and on a need for European competitors for existing services provided by

players from other continents. The **public broadcaster archives** and the **public broadcaster research organisation** mention a need for technologies that understand, classify and sort audiovisual material by analysing sound, images, video and text automatically. They also mention a need for Linked Data technologies to intelligently link metadata throughout the production process. The automated extraction of metadata, preferably on a scene level, using analytics tools is mentioned by two stakeholders, i.e. a need for better indexing and retrieval possibilities. The financing network indicates a need for AI technologies that help predict how a certain film, television series or video game will perform in the market. The news agency has a need for smart tools that recommend synonyms, syntactic constructions, certain words etc. when writing a piece of text content. For images, they have a need for verification methods and for metadata, a need for automated enrichment technologies. The cultural organisation suggests creating a truly European digital public space using LT and MT.

Q4: EU policies – Do you see a need for establishing EU policies (including regulation and recommendations) for the use of AI-based technologies in the audiovisual sector and why?

In terms of EU policies and regulation, most **technology and service providers** suggest concentrating on an ethical framework regarding the use and misuse of AI (ELSI, ethical, legal, security implications) that protects human values and equal opportunities, and fosters cultural and linguistic diversity. EU policies should be put in place to protect against the misuse of AI for textual and audiovisual false news and misinformation and also against the misuse of AI in commercial sales, i.e. restraining persuasive AI-driven ecommerce. This also relates to the definition of basic requirements with respect to training data and trained models, especially in the area of bias and misinformation, for which keeping track of the provenance of data and content plays a significant role, for example, as a type of veracity stamp that prevents forgery and misuse (deep fakes, voice cloning etc.). However, the respondents also emphasise that regulation should not hinder innovation. With regard to the dominance of very large enterprises from other continents, the respondents suggest to support alternative technologies provided by European companies, and that European companies should have the same level of access to European data as the big technology players from the US and Asia. The stakeholders also suggest establishing guidelines and policies both for AI technologies themselves and for their aims and purposes including data source selection, data sharing, data holding as well as algorithm and service providers. They also mention a need for consistent recommendations for closed captioning to improve the accessibility of content (including standards). Other topics emphasised are better EU regulation regarding copyright laws, IPR, data management and competition, sustainable AI, trustworthy AI and explainable AI as well as common data formats and standards for taxonomies that describe which types of content have been annotated by AI tools.

Q5: Opportunities of AI in the content life-cycle – AI-based technologies can be used for the creation, production or consumption of audiovisual content. In which of these three broader phases of the content life-cycle do you see the biggest opportunities for AI-based technologies and why?

AI creates many new opportunities in the content life-cycle. Strong feedback was received regarding all three phases, multiple respondents mentioned the need to focus on high quality content. The **technology and service providers** perceive opportunities especially in the consumption phase, because it allows for automated recommendations, discovery, search and translation of content (cross-language access). Regarding the creation and production phases, applications such as the AI-based generation of soundtracks, automated subtitling and transcription as well as the creation of 3D characters, character movement, conversations and interactions were mentioned (including special effects in film and television programmes). One respondent speculates that the production and consumption phases will sooner or later merge into one phase within a personalised and data-driven ecosystem. The **public broadcasters** also emphasise production and consumption but mention that at least part of the creation phase is

addressed by automated text generation (Natural Language Generation, NLG), offered by service providers. NLG technologies are able to generate weather or stock exchange reports or articles about football matches, for example. The public broadcasters perceive AI as a crucial technology for supporting human experts in the process of editing, producing and distributing content. It would allow for processes that would not be possible otherwise. Furthermore, NLG and related technologies can be used to amplify the amount of content produced, including for experimental purposes such as hyper local television programmes. The consumption phase was also interpreted as *automated* consumption, i.e. the monitoring of media channels for content verification and analytics purposes, for example, to identify trends in competitors' content or controlling internal content cross-lingually. The **public broadcaster archives** emphasise the preservation and re-use of audiovisual content as a huge opportunity, while the **research** organisations focus on a more digital AI-driven content production, which would result in high quality content and also fine-grained metadata descriptions. Other stakeholders mention the opportunities of AI-based tools especially for the post-production phase (content localisation using machine translation, video tagging using object and scene detection, metadata creation using text analytics etc.). Opportunities were also mentioned for the book publishing sector regarding the improvement of processes in publishing houses and better access to content on the side of the customer (accessibility, content discovery, copyright management etc.).

Q6: Dangers of AI in the content life-cycle – AI-based technologies can be used for the creation, production or consumption of audiovisual content. In which of these three broader phases of the content life-cycle do you see the biggest danger of AI-based technologies and why?

Regarding the dangers of AI use in the content life-cycle, many of the 22 respondents mention the area of media consumption with intentionally created false content, biased content, manipulative content and deep fakes. AI-based technologies can be used to create such content, to shape and also to distribute it according to certain demographic data (collected via AI-based technologies) while the large AI-based social media platforms function as amplifiers of false content – corresponding detection and verification methods are needed as well as methods for presenting that information to the user in a transparent way. In addition, **technology and service providers** emphasise the large-scale, AI-based monitoring of EU citizens (via social media networks) as a concrete danger. In terms of media consumption, it is well known that all platforms attempt to keep users on their platforms as long as possible through AI-driven recommendations. In that regard, users need to be provided with guidelines and educational tools to prevent addictive and also manipulative media consumption habits – this could perhaps be implemented as part of self-regulation. At the same time, there is a danger that recommendation engines may not be able to classify and to suggest all types of content, which in turn creates the risk for large amounts of high-quality content to be lost forever because they will never be recommended to the intended target audiences. Furthermore, for the phase of media creation and production, the **public broadcaster** stakeholders mention the danger that AI-based technologies, if not supervised and used properly, could lead to an influx of uncontrolled, low-quality and untrustworthy content. Another relevant topic is data sovereignty, i.e. when data is handed over to a service provider for a specific purpose, data should remain under the control of its owner. In terms of additional dangers mentioned, there is a risk that AI technologies are not affordable or that organisations end up in company-specific AI silos around which they develop their IT infrastructure (vendor lock-in). IPR issues related to the input and output of AI need to be addressed as well. Finally, there are many workflows that involve both human experts and AI-based technologies, for example, in the localisation area. Here, training and remuneration for the staff involved need to be put in place to avoid low quality content.

Q7: European AI tool market – Please provide a short description of how you assess the market of European AI tool providers for the audiovisual sector (for example, in terms of market access, market size, market diversity, pricing, competitiveness, market dominance etc.).

The European AI tool market is perceived by the respondents in a highly interesting way. The **technology and service providers** state that large US companies dominate the market (and should be taxed properly by Europe, according to one respondent), which is why Europe should support and interconnect European AI companies and research. In that regard, big US technology enterprises were mentioned as well as media and entertainment companies like Netflix and Amazon Prime. According to one respondent, the best positioned European stakeholders seem to be the platforms put in place by telecom operators like Telefónica and Orange. Some respondents perceive the European market with genuinely European providers as non-existent. Two others state that the European landscape of providers of AI and LT tools is very broad and extremely fragmented, especially at a national, i.e. language-related level. There are numerous companies, mostly SMEs and startups, that provide tools and services for the audiovisual and media sector. However, most of these tools are assessed as being too generic: they lack a proper integration layer for developing efficient processing workflows. Most of these SMEs do not have the ability to scale but, nevertheless, they provide competitive, high quality services for reasonable prices. As a consequence, the market is immature and disconnected. The research community does not seem to support systematic progress as much as it could. In addition, some of the stakeholders from the **public broadcaster** group indicate that they have started to develop their own AI-based technologies in European projects, often with the help of research organisations and SMEs (see above).

Q8: Opportunities of AI (in general) – Independent of your organisation, in which areas do you see the biggest potential or set of opportunities (including technical, social, ethical, economic, legal aspects) for AI-based technologies in the audiovisual sector?

The respondents mention various opportunities in addition to those addressed in Q5. The **technology and service providers** emphasise opportunities for content creation and production. In these, AI allows unlimited localisation and makes it possible for a fragmented and culturally diverse ecosystem to survive in a world that is currently dominated by capital-intensive ventures based in the US. One goal could be to industrialise the process of media production, making high-end technology available to mid-size producers so they can make a better product faster and for less money. Specifically, the respondents mention translation tools, search engines, content recommendation engines and content curation tools – not only for knowledge workers but also for citizens so that they can better handle the flood of data, which is continuously increasing. Another crucial opportunity mentioned is the creation of a European Digital Public Space, i.e. a European Media Platform with components for cross-lingual and interactive communication that cannot be manipulated through intentionally created misinformation. In that regard, AI can support European society through services that prevent false news and information bias and that improve accessibility, inclusiveness and integration, helping the European society to grow into a direction that is the most beneficial in socio-cultural terms, respecting European values. The stakeholders from the **public broadcaster** group emphasise the opportunities created by AI-based tools that help free up valuable resources, i.e. tools that enable human experts to do their day-to-day work in a more efficient and better way (and also to delegate uninteresting routine tasks to the machine). There are also opportunities for novel kinds of business models for the distribution of content, for example, regarding payments based on the size of the audience reached. The use of deep fakes in a legal and creative way could enhance the production value and, thus, act as an efficiency driver. This also relates to the use of AI for special effects and AR/VR productions, i.e. the democratisation of special effects. Several stakeholders emphasise opportunities of AI-based technologies to improve the accessibility of audiovisual content and media for people with disabilities, ethnic minorities or visitors from other countries, for example, through automated subtitling and transcription, machine translation, automated image or scene descriptions, automated tagging etc. The last area mentioned relates to archives, especially the digitisation of cultural heritage (old books, documents etc.).

Q9: Dangers of AI (in general) – In which areas do you see the biggest dangers or threats (including technical, social, ethical, economic, legal aspects) of AI-based technologies in the audiovisual sector?

In addition and on top of the feedback provided regarding Q6 (especially fake news, false news, deep fakes, voice cloning, disinformation campaigns, social and ethical aspects), the respondents stress a number of dangers of AI-based technologies in the audiovisual sector. As video is quickly becoming the main means of communication in our societies, there is a severe threat in relation to the use of AI-based technologies specifically concerning misinformation and manipulation, which could severely impact the foundation of our democratic society. There is also a sustainability issue vis-a-vis the energy consumption required for these technologies to run and the impact their increasing use has on the environment. Another threat relates to the fear that AI could be concentrated in the hands of a few global players, mainly companies or governments, and be (mis)used for their personal interests. This danger can only be addressed through transparent and explainable AI, real competition and an adequate legal framework. A related threat involves the danger of a "big brother" scenario, in which the complete surveillance of citizens is performed by some governments. As another respondent notes, it is important to consider the risks and address them in advance or parallel to their actual development. Otherwise Europe runs the risk that the pace of AI development constantly outpaces legislation on ethics and markets.

Q10: European collaboration – In which broader areas regarding the use of AI-based technologies in the audiovisual sector do you see a need for increased European collaboration or coordination?

Most responses in terms of increased European collaboration have already been discussed in previous questions. The respondents perceive collaboration as essential because individual players have limitations and difficulties to use AI-based technologies (lack of resources and know-how). Collaborations are suggested between players along the media production cycle (analysis, creation, distribution, feedback) or in a cross-sectorial way, organised, funded and coordinated by the EU. In that regard, a European platform for AI applications could help, also to publicly trade training data, language models etc. Many respondents mention data sets as an important aspect of collaborations, for example, the joint creation and annotation of open data sets or the joint development of open source tools to be used in research and exploitation or for benchmarking purposes. Europe's multilingualism was mentioned several times as a crucial collaboration topic. For example, to guarantee inclusiveness and accessibility through closed captioning and multilingual technologies, corresponding tools need to be made available, especially for under-resourced languages. As the European Language Technology landscape is fragmented, coordination and collaboration is needed among the various players, most of which are SMEs, in order to concentrate efforts and to stimulate the market through joint and shared grid platforms. In addition to the suggestion to increase AI research and standardisation efforts in Europe, one respondent suggests to collaborate in the area of digital forensics (video, audio, photo, text), content verification and disinformation analysis. The Audiovisual Media Services Directive (AVMSD) was not mentioned in any of the responses.

Survey: Main Findings

Q1: Current use of AI – Broad production use and partial experimental use of AI technologies in various stakeholders of the audiovisual sector including public broadcasters, especially for automated indexing (metadata generation), improved accessibility and media as well as content localisation. AI is used for processing audio or video data, language or text data or for information and knowledge management purposes. Among the technologies used are ASR, TTS, NLP, NER, MT, summarisation, search and recommender engines, content classification, automatic subtitling, computer vision methods (object detection, face detection etc.) and metadata management.

Q2: Future use of AI – In the future, the AI technologies mentioned above will be improved. They will be deployed in broader use cases in the next 24 months. AI technologies foreseen for future use are slightly more avantgarde and not as mature as standard off-the-shelf technologies such as ASR, TTS and MT. Examples include automated detection of illegal content, deep fakes, synthetic content and flexible curation technologies that include chatbots or question answering. Furthermore, more experimental use of AI technologies is foreseen by the more conservative stakeholders.

Q3: Needs, demands and requirements – All stakeholders have a demand for large amounts and different types of training data including labelled structured and unstructured data, domain-specific training data, acoustic data and data for illegal content. There is also a need for more language technologies with a broader coverage (also including European languages with smaller numbers of speakers) including ASR, TTS, MT and content curation services. A demand was voiced for semantic vocabularies that can be used for metadata extraction and enrichment as well as Linked Data technologies to intelligently link metadata throughout the production process. There is a need for European competitors for existing services provided by players from other continents.

Q4: EU policies – Most stakeholders suggest focusing on an ethical framework regarding the use and misuse of AI that protects human values and fosters cultural and linguistic diversity. Policies should protect against the misuse of AI for false news and misinformation and also against the misuse of AI in commercial sales. The respondents also emphasise, though, that regulation should not hinder innovation and that European companies should be supported more because of the dominance of the big technology enterprises from other continents. Other topics are better EU copyright laws, data management and competition, sustainable, trustworthy and explainable AI.

Q5: Opportunities of AI in the content life-cycle – Many respondents suggest concentrating on the production of high quality content. Opportunities are also perceived in the consumption phase because it allows for recommendations, discovery, search and content translation. For creation and production, AI-based generation of soundtracks, automated subtitling and transcription as well as 3D characters and special effects are mentioned. Stakeholders perceive AI to be a crucial technology to support human experts in the process of editing, producing and distributing content. The automated production, post-production and also preservation and re-use of existing audiovisual content is seen as a huge opportunity.

Q6: Dangers of AI in the content life-cycle – Many respondents mention media consumption with intentionally created false or manipulative content and deep fakes as an imminent danger. The large-scale, AI-based monitoring of EU citizens is another danger. In terms of media consumption, all platforms attempt to keep users on their platforms as long as possible through AI-driven recommendations. Users need to be provided with guidelines and tools to prevent addictive and manipulative media consumption habits. For creation and production, stakeholders mention the danger that AI tools could lead to masses of uncontrolled, low-quality and untrustworthy content.

Q7: European AI tool market – The European AI tool market is not perceived by all respondents in the same way. Some state that large US companies dominate the market, which is why the EU should support European AI companies and research. Others perceive the European market with European providers as non-existent. Yet others state that the European landscape of providers of AI and LT tools is broad yet extremely fragmented, especially on a national, i.e. language-specific level, with numerous companies, mostly SMEs and startups, that provide tools and services for the audiovisual and media sector. A few stakeholders have started to develop their own AI-based technologies in European projects, often with the help of research organisations and SMEs.

Q8: Opportunities of AI (in general) – The respondents mention various opportunities in addition to those addressed in Q5. For content production, AI enables unlimited localisation and makes it possible for a fragmented and culturally diverse ecosystem to survive in a world

dominated by capital-intensive ventures based in the US. Further, it is suggested to industrialise media production, making high-end technology available to mid-size producers. Another opportunity is a large European Digital Public Space, i.e. a European Media Platform with cross-lingual and interactive communication tools that cannot be manipulated through misinformation. Other stakeholders emphasise the opportunities created by AI tools that help free up valuable resources. Several stakeholders mention opportunities of AI technologies to improve the accessibility of audiovisual content and media for people with disabilities, ethnic minorities or visitors from other countries.

Q9: Dangers of AI (in general) – On top of the feedback to Q6, the respondents stress that, as video is quickly becoming our main means of communication, there is a threat that relates to the use of AI for misinformation and manipulation, which could have a severe impact on the foundations of democratic societies. There is also a sustainability issue (energy consumption, impact on the environment). Another threat relates to the fear that AI could be concentrated in the hands of a few global players and be (mis)used for their personal interests. A related threat involves the danger of a "big brother" scenario, in which the surveillance of citizens is performed by some governments.

Q10: European collaboration – Collaboration is seen as essential because individual players have limitations and difficulties to use AI technologies. Collaborations are suggested along the media production cycle or in a cross-sectorial way. To this end, a European platform for AI applications could help. Data sets are seen as important, for example, the joint creation of open data sets or the joint development of open source tools. Europe's multilingualism is seen as crucial: to guarantee inclusiveness and accessibility through closed captioning and multilingual technologies, tools need to be made available, especially for under-resourced languages. As the European Language Technology landscape is fragmented, coordination and collaboration is needed in order to concentrate efforts through joint and shared grid platforms. It was also suggested to collaborate in digital forensics (video, audio, photo, text), content verification and disinformation detection and analysis.

Summary of Key Survey Responses with a Focus on Q1, Q2 and Q3

	Production Use	Experimental Use, Future Use, Interesting Novel Areas	Important Needs
Audio	- Transcription of audio (ASR) for closed captioning and subtitling (improved accessibility) - Speech generation (TTS)	- Speaker classification - Emotion recognition - Acoustic scene detection - Automated detection of illegal content (classification) - Automated detection of deep fakes and synthetic media (classification)	- More annotated structured and unstructured data for training AI systems - More and better data, services and technologies for more EU languages - Multimedia forensics and content verification: detection of false news, misinformation etc. - Transparent and trustworthy AI
Image and Video	- Transcription of audio (ASR) for closed captioning and subtitling (for improved accessibility) - Speech generation (TTS) - Image classification (for indexing) - Object detection (for indexing) - Face detection (for indexing) - Scene detection (for indexing) - Automatic rough cuts - Raw footage processing - AI to enhance video encoding and content delivery	- Automated detection of illegal content (classification) - Automated detection of deep fakes and synthetic media (classification) - Predicting the location of photos	- More annotated structured and unstructured data for training AI systems - Multimedia forensics and content verification: detection of false news, misinformation etc. - Transparent and trustworthy AI
Text	- Translation of text content (MT) (for improved accessibility) - Named Entity Recognition (NER) (for	- Automated detection of illegal content (classification) - Automated detection of deep	- More annotated structured and unstructured data for training AI systems

	indexing) - Summarisation (for indexing) - OCR (for indexing) - Concept and meaning extraction (for indexing) - Internal search engines - Recommender engines - Content classification (for indexing) - Topic detection (for indexing) - Automated subtitling - Localisation - Social Media Analytics and Monitoring	fakes and synthetic media (classification) - Chatbots (AI for better support workflows) - Question Answering (AI for better support workflows)	- More domain-specific terminologies and semantic vocabularies for metadata enrichment - More and better data, services and technologies for more EU languages technologies that support content creation and content curation - Multimedia forensics and content verification: detection of false news, misinformation etc. - Transparent and trustworthy AI
Information and Knowledge Management	- Management of metadata that were automatically extracted using the AI technologies mentioned above (indexing) - Knowledge Graphs and Linked Data platforms for linking semantic metadata		- More domain-specific terminologies and semantic vocabularies for metadata enrichment - Standardised semantic taxonomies for annotation - Better tools for metadata linking in the whole production process

Concluding Remarks and Recommendations

The recent breakthroughs in AI research and overall popularity of the topic have acted as a key driver of the already existing convergence of modalities (audio, image, video, text/language and also interactive speech devices and technologies). While "multimedia" applications were somewhere between research and initial core seeds in the 1990s (especially in the form of CDROMs and the World Wide Web), such integrated and multimodal, often even interactive, media products are now ubiqutous in our multi-device landscape (smart television sets, tablets, mobile phones, laptops, smart watches, car dashboards, conversational agents, smart household appliances etc.). Among others, AI is contributing to the creation, planning, managing, production, distribution, localisation and consumption of audiovisual media products.

With regard to Europe, one overarching and inherent dimension that relates to all audiovisual, multimedia and multimodal products and productions is language, either in spoken or in written form. Europe is a multilingual society with 24 official EU Member State languages and approx. 60 additional languages (unofficial languages, minority languages, languages of immigrants and important trade partners). According to the Treaty of the European Union, all languages are considered equal. Yet, support for the different European languages through technologies is highly imbalanced with English being the best supported language by far, as the large-scale study "Europe's languages in the digital age", published by the EU Network of Excellence META-NET in 2012, has found. Since then, the situation regarding technology support for Europe's languages has improved but it is still considerably imbalanced. These Language Technologies (LT) – also called language-centric AI technologies – are at the core of the survey findings, summarised in the previous section. One of the most recent studies and reports that demand balanced technology support for our languages is the European Parliament report "Language equality in the digital age", jointly prepared by the CULT and ITRE committees in 2018. The report received broad support in the EP on 11 September 2018 with a landslide 592 votes in favour of the report.

While LT is an inherent sub-field of AI, the field of AI at large has also seen a lot of activity in recent years, not only in Europe but also globally. Both the European LT landscape and also the general European AI landscape are very much fragmented, especially in terms of their respective markets and companies active in these markets. Crucially, this also includes the market of AI and LT tools and services that is in scope of the organisations that create, produce and distribute audiovisual media and products such as, among others, public and private broadcasters. In the European LT

landscape and also in the survey conducted for this briefing paper, we have seen, time and again, very little market awareness, i.e. potential buyers/purchasers of LT or AI technologies are, simply, *not* aware of European offerings in this area. On top of this situation, the respective markets themselves, especially the LT market, are populated by a very large number of SME companies that focus on their local and regional markets and customers. In the LT landscape, only a small number of bigger companies exist that are currently attempting to scale up.

Due to the general landscape and market situation, two European projects currently develop umbrella platforms with the main goal of bringing all these distributed and fragmented players together under a joint and shared roof. The project AI4EU currently develops the "European AI on Demand platform", which is supposed to include all European players, technologies, services and data sets. The project European Language Project (ELG) develops a similar platform but concentrates on Language Technologies. The ELG platform will ultimately include all European commercial and research organisations active in the field as well as functional services, data sets, catalogues of stakeholders. The results of the survey show that these umbrella platforms are being developed at the right time, given that some of the survey respondents appear to have little market awareness.

This briefing paper concludes with four recommendations that were derived from and informed by the survey summarised in the first half of the briefing paper.

Recommendation 1: Support Long-Term European Initiatives

The survey responses clearly demonstrate a need for overarching umbrella platforms for different types of AI technologies, especially general AI tools and also Language Technologies. Two umbrella platforms, AI4EU and ELG, have been under development since the beginning of 2019. First versions of both platforms were launched in early 2020. Both projects collaborate on the challenging topic of platform interoperability, i.e. they want to make their platforms aware of each other so that, for example, a search for a certain tool in AI4EU also delivers results from ELG and vice versa.

The first recommendation is to continue the support of and, in fact, reinforce the relevant and emerging platforms and long-term initiatives such as, crucially AI4EU and ELG. One of the key reasons why these domain-agnostic platforms are being developed is to provide improved access to the whole AI and LT tools market including commercial and academic organisations and repositories of tools and data sets. These emerging platform initiatives need to be strengthened.

In addition to AI4EU and ELG, a new relevant project will start at the end of 2020: AI4MEDIA. This new project is funded through Horizon 2020 call ICT-48-2020 and will specifically focus upon AI technologies for the media sector. AI4MEDIA will collaborate with the wider "European AI on demand platform" ecosystem of platforms.

Recommendation 2: Reinforce European AI & LT Provider Landscape

The European AI and LT tools and service provider landscape is, for the most part, world-class and dominated by many SMEs. Unfortunately, like many other European markets, the landscape and market is characterised by a high level of fragmentation, which results in little market awareness on the side of potential customers and clients. On top of this situation, the large technology enterprises from the US and, to a certain degree, also Asia, dominate many areas of IT and ICT

including the audiovisual sector.[1] Many European offerings exist as well, but they are more fragmented and, hence, not as feature-rich, thus requiring more effort regarding their integration into existing internal platforms and infrastructures.

The second recommendation is to significantly reinforce the European AI and LT tools and service provider landscape through funds for targeted support actions, collaborative research projects, experimental pilot development, joint data set annotation and distribution, open source tool development, foster and boost European standardisation activities etc. The European audiovisual sector and media landscape would benefit immensely from more intense collaboration in this area, as also expressed by many of the survey respondents. One possible strategic umbrella under which to orchestrate this increased collaboration is described in the third recommendation. Corresponding funds could be provided through Digital Europe, Creative Europe and Horizon Europe.

Recommendation 3: Support the European Digital Public Space Vision (European Media Platform)

According to many authors, observers and politicians, one of the most crucial gaps of the European project is the lack of a European Public Space in which all European citizens can inform themselves about current European developments and in which they can discuss important topics and share their viewpoints on a pan-European level.[2] With the help of digital technologies, it is now finally possible to develop and to put into place a European Digital Public Space, i.e. to fill this crucial gap using a digital media and communication platform that serves all of Europe. Broadcasters, companies, research organisations, NGOs and all other stakeholders including citizens could share all types of media content (from television programmes through podcasts to short user-generated videos), all citizens could come together to discuss important topics and share their opinions. A crucial aspect of such a platform is that it must not be vulnerable to external manipulation attempts, for example, through the same type of orchestrated disinformation campaigns that have been eroding the foundations of our democratic society. In terms of technical prerequisites, such a platform would need a diverse set of AI technologies for all audiovisual aspects including Language Technologies to enable all European citizens not only to use the platform in their own mother tongue but also to participate in online discussions with other European citizens in their own mother tongues. Machine translation and related multilingual and cross-lingual technologies would need to be put in place to enable, for the very first time, communication across language barriers on such a large scale, i.e. on the scale of a whole continent. ASR is needed for automatic indexing, metadata generation and automatic subtitling. MT can be used to translate automatically generated subtitles on the fly. Even automatic speech-to-speech translation could be foreseen for the near future. Face and object detection can be used for metadata generation and indexing. This is just a brief list of selected AI-driven technologies – many additional ones can be foreseen.

The third recommendation is for the European Parliament to support the development of such a European Digital Public Space, also called "European Media Platform" by some initiatives. This vision would be an ideal umbrella initiative and vision to put in place coordination and support activities for the audiovisual sector on a long term basis, to bring together the various fragmented

[1] For example, one technical framework that is very popular with many public and private broadcasters is the Microsoft Azure Media Services, see https://azure.microsoft.com/en-us/services/media-services/. The connection of this framework to the Microsoft Azure cloud product and also its broad feature set make its integration into existing IT platforms and internal infrastructures quite simple.

[2] See, for example, „Hier spricht Europa", Jakob von Weizsäcker and André Wilkens, 10 December 2016. https://www.spiegel.de/politik/ausland/europafunk-gegen-die-propaganda-flut-essay-a-1123991.html

communities and stakeholders, to initiate collaborations between, among others, public broadcasters and research centres and AI/LT tool and service providers. Such an activity would require an enormous, coordinated development programme, almost like an Airbus programme for the European media and IT sector.

Fortunately, there are already several initiatives that are trying to help kickstart such a large programme. One of them was started by Ulrich Wilhelm, Director of Bayerischer Rundfunk (BR), the public broadcaster in Germany's federal state of Bavaria.[3] Wilhelm is positioning such a European Media Platform as the European answer to Facebook, Google and Netflix. Currently, an initial concept and feasibility study is prepared by one of the German Federal Ministries involved in the discussion. It can almost be said that the development of such a platform is mission-critical for the further development of the European continent, which is why broad support is to be expected. In addition to this initiative, there are several others. For example, recent recommendations published by the European Broadcasting Union (EBU) go into the same direction.[4]

Recommendation 4: Conduct In-Depth Study of the Sector

The last recommendation refers to the fact that the survey presented in this briefing paper has only been able to scratch the surface. It was organised, conducted and analysed under immense time pressure, which is why its findings can only be considered indicative. A similar survey should be conducted on a much bigger scale, covering the whole of Europe with a clearer focus on key questions. The findings of the present survey can be used to provide a better focus and also seed information, for example, regarding multiple-choice questions such as types of AI technologies already in use. The result of such a broader survey should be a clear list of the concrete needs and wishes of all European stakeholders regarding AI technologies in the audiovisual sector. Such a study could be conducted in collaboration with one of the EU-funded projects in this domain.

Appendix 1: A Selection of Relevant Projects and Initiatives

Various EU-funded projects and initiatives work on closely related topics and challenges. A selection of relevant projects and initiatives is included below. This list is not meant to be exhaustive.

- **European Language Grid** (**ELG**) is a Horizon 2020-supported three year project that currently develops the primary Language Technology platform for Europe. ELG will eventually host thousands of data sets and hundreds of functional LT/NLP/Speech services. As the European Language Technology landscape is highly fragmented, ELG wants to provide a joint umbrella platform for all stakeholders that belong to this community, providing reach and visibility. https://www.european-language-grid.eu
- **AI4EU** is a Horizon 2020-supported three year project that currently develops the European AI on demand platform. ELG (see above) and AI4EU collaborate closely, especially on the topic of platform interoperability. https://www.ai4eu.eu
- **META-NET** is a European Network of Excellence that consists of 60 research centres in 34 European countries. META-NET works on the technological foundations of the multilingual European information society. http://www.meta-net.eu

[3] See „ARD-Vorsitzender Ulrich Wilhelm: Europäische Medienplattform "ein sehr dickes Brett"", https://meedia.de/2019/12/13/ard-vorsitzender-ulrich-wilhelm-europaeische-medienplattform-ein-sehr-dickes-brett/

[4] https://www.ebu.ch/news/2019/10/european-media-innovation-key-to-competitiveness

- **NEM** (New European Media) is the leading European Network for Media and the Creative Industries. Its mission is to foster the impact of interactive technologies on the future of new media. https://nem-initiative.org
- **MeMAD** (Methods for Managing Audiovisual Data) is a Horizon 2020-supported project that investigates novel methods for accessing and using audiovisual content including several of the topics mentioned in this briefing paper. https://memad.eu
- **GoURMET** (Global Under-Resourced Media Translation) is a Horizon 2020-support project that develops neural machine translation technologies as well as tools for media analysts and journalists. https://gourmet-project.eu
- **SUMMA** (Scalable Understanding of Multilingual Media) was a Horizon 2020-supported project that integrated media processing tools (including ASR, MT) with deep language understanding capabilities, implemented in media use cases. http://summa-project.eu

Appendix 2: Glossary – Terms and Abbreviations[5]

Artificial Intelligence (AI) – In computer science, artificial intelligence (AI) is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. Colloquially, "AI" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".

Application Programming Interface (API) – An API is a computing interface to a software component or a system, that defines how other components or systems can use it. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc. It can also provide extension mechanisms. An API can be entirely custom, specific to a component, or it can be designed based on a standard to ensure interoperability. Since other components/systems rely only on the API, the system that provides the API can (ideally) change its internal details "behind" that API without affecting its users. Today, with the rise of REST and web services over HTTP, the term is often assumed to refer to APIs of such services.

Audiovisual (AV) – Audiovisual (AV) is digital media possessing both a sound and a visual component, such as films, television programs, multimedia content (including text and hypertext, especially on the World Wide Web) and corporate conferencing.

Augmented Reality (AR) – Augmented reality (AR) is an interactive experience of a real-world environment where the objects that reside in the real world are enhanced by computer-generated perceptual information, sometimes across multiple sensory modalities. AR can be defined as a system that fulfills three basic features: a combination of real and virtual worlds, real-time interaction, and accurate 3D registration of virtual and real objects. The overlaid sensory information can be constructive (i.e. additive to the natural environment), or destructive (i.e. masking of the natural environment). This experience is seamlessly interwoven with the physical world such that it is perceived as an immersive aspect of the real environment.

Automatic Speech Recognition (ASR) – Speech recognition is an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. It incorporates knowledge and research in computer science, (computational) linguistics and

[5] Most explanations in this glossary are based on definitions taken from the corresponding Wikipedia pages.

computer engineering. Speech recognition applications include voice user interfaces such as voice dialing (e.g. "call home"), call routing (e.g. "I would like to make a collect call"), search key words (e.g. find a podcast where particular words were spoken), simple data entry (e.g. entering a credit card number), preparation of structured documents (e.g a radiology report), determining speaker characteristics, speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed direct voice input). Speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced by the surge of academic papers published in the field and by the worldwide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems.

Closed Captioning (CC) – Closed captioning (CC) and subtitling are both processes of displaying text on a television, video screen, or other visual display to provide additional or interpretive information. Both are typically used as a transcription of the audio portion of a program as it occurs (either verbatim or in edited form), sometimes including descriptions of non-speech elements. HTML5 defines subtitles as a "transcription or translation of the dialogue when sound is available but not understood" by the viewer (for example, dialogue in a foreign language) and captions as a "transcription or translation of the dialogue, sound effects, relevant musical cues, and other relevant audio information when sound is unavailable or not clearly audible". Improvements in speech recognition technology means that live captioning may be fully or partially automated. BBC Sport broadcasts use a "respeaker": a trained human who repeats the running commentary (with careful enunciation and some simplification and markup) for input to the automated text generation system. This is generally reliable, though errors are not unknown. Recent breakthroughs in automatic speech recognition have lead to better and more robust ways of producing closed captions and subtitles for a wider range of spoken language embedded in a video.

Curation Technology – Digital content and online media have gained immense importance, especially in business, in industry, in the audiovisual sector, but also in politics and other areas of society. Some of the challenges include better support and smarter technologies for digital content curators who are exposed to an ever increasing stream of heterogeneous information they need to process further. For example, professionals in a digital agency create websites or mobile apps for customers who provide documents, data, pictures, videos etc., that are processed and then deployed as new websites or mobile apps. Knowledge workers in libraries digitize archives, add metadata and publish them online. Journalists need to continuously stay up to date to be able to write a new article on a specific topic. Many more examples exist in various industries and media sectors (television, radio, blogs, journalism, etc.). These diverse work environments can benefit immensely from smart AI-based semantic content and media curation technologies that help content curators, who are usually under great time pressure, to support their processes.

Deep Learning (DL) – Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

Hypertext Transfer Protocol (HTTP) – HTTP is an application protocol for distributed, collaborative, hypermedia information systems. HTTP is the foundation of data communication for the World Wide Web, where hypertext documents include hyperlinks to other resources that the user can easily access, for example by a mouse click or by tapping the screen in a web browser.

Development of HTTP was initiated by Tim Berners-Lee at CERN in 1989. Development of early HTTP Requests for Comments (RFCs) was a coordinated effort by the Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C), with work later moving to the IETF.

Information and Communication Technology (ICT) – Information and communications technology (ICT) is an extensional term for information technology (IT) that stresses the role of unified communications and the integration of telecommunications (telephone lines and wireless signals) and computers, as well as necessary enterprise software, middleware, storage, and audiovisual systems, that enable users to access, store, transmit, and manipulate information. The term ICT is also used to refer to the convergence of audiovisual and telephone networks with computer networks through a single cabling or link system.

Intellectual Property Rights (IPR) – Intellectual property rights include patents, copyright, industrial design rights, trademarks, plant variety rights, trade dress, geographical indications, and in some jurisdictions trade secrets.

Information Technology (IT) – Information technology (IT) is the use of computers to store, retrieve, transmit, and manipulate data or information. IT is typically used within the context of business operations as opposed to personal or entertainment technologies. IT is considered to be a subset of information and communications technology (ICT).

Linked Data (LD) – In computing, Linked Data is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read by computers. Tim Berners-Lee, director of the World Wide Web Consortium (W3C), coined the term in a 2006 design note about the Semantic Web project. Linked data may also be open data, in which case it is usually described as linked open data (LOD).

Language Technology (LT) – Language technology, often called human language technology (HLT), studies methods of how computer programs or electronic devices can analyze, produce, modify or respond to human texts and speech. It consists of natural language processing (NLP) and computational linguistics (CL) on the one hand, and speech technology on the other. It also includes many application oriented aspects of these. Working with language technology often requires broad knowledge not only about linguistics but also about computer science.

Knowledge Graph (KG) – In computer science and information science, a knowledge graph (also known as an ontology) encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse. A knowledge graph is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject. One common interpretation is that a knowledge graph represents a collection of interlinked descriptions of entities – real-world objects, events, situations or abstract concepts. Knowledge graphs often contain large volumes of factual information.

Metadata – Metadata is data that provides information about other data. In other words, it is data about data. Many distinct types of metadata exist, including descriptive metadata, structural metadata, administrative metadata, reference metadata and statistical metadata. In the audiovisual and media sector, it is especially descriptive metadata that is relevant as descriptive information about a resource. It is used for discovery (search) and identification. It includes elements such as title, abstract, author, and keywords. Often, automated methods help with the extraction of descriptive metadata from media sources including video, audio and text.

Machine Learning (ML) – Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks.

Machine Translation (MT) – Machine translation, sometimes referred to by the abbreviation MT (not to be confused with computer-aided translation, machine-aided human translation or interactive translation), is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another. The best performing methods are based on neural technologies (Deep Learning).

Multimedia – Multimedia is content that uses a combination of different content forms such as text, audio, images, animations, video and interactive content. Multimedia contrasts with media that use only rudimentary computer displays such as text-only or traditional forms of printed or hand-produced material.

Named Entity Recognition (NER) – Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc. NER is often based on neural technologies (Deep Learning).

Natural Language Generation (NLG) – Natural-language generation (NLG) is a software process that transforms structured data into natural language. It can be used to produce long form content for organizations to automate custom reports, as well as produce custom content for a web or mobile application. It can also be used to generate short blurbs of text in interactive conversations (a chatbot) which might even be read out by a text-to-speech system.

Natural Language Processing (NLP) – Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. Challenges in NLP frequently involve speech recognition, natural language understanding, and NLG.

Optical Character Recognition (OCR) – Optical character recognition is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast). Widely used as a form of data entry from printed paper data records – whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation – it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

Representation State Transfer (REST) – Representational state transfer (REST) is a software architectural style that defines a set of constraints to be used for creating Web services. Web services that conform to the REST architectural style, called RESTful Web services, provide

interoperability between computer systems on the Internet. RESTful Web services allow the requesting systems to access and manipulate textual representations of Web resources by using a uniform and predefined set of stateless operations.

Speech Synthesis – see Text to Speech (TTS)

Speech Technology – Speech technology relates to the technologies designed to duplicate and respond to the human voice. They have many uses. These include aid to the voice-disabled, the hearing-disabled, and the blind, along with communication with computers without a keyboard. The subject includes several subfields: Speech synthesis, Speech recognition (ASR), Speaker recognition, Speaker verification, Speech encoding, Multimodal interaction.

Subtitling – see Closed Captioning (CC)

(Automatic) Summarisation – Automatic summarization is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content. In addition to text, images and videos can also be summarized. Text summarization finds the most informative sentences in a document; image summarization finds the most representative images within an image collection; video summarization extracts the most important frames from the video content.

Text to Speech (TTS) – Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.

Virtual Reality (VR) – Virtual reality (VR) is a simulated experience that can be similar to or completely different from the real world. Applications of virtual reality can include entertainment (i.e. video games) and educational purposes (i.e. medical or military training). Other, distinct types of VR style technology include augmented reality and mixed reality.

World Wide Web (WWW) – The World Wide Web (WWW), commonly known as the Web, is an information system where documents and other web resources are identified by Uniform Resource Locators (URLs, such as https://www.example.com), which may be interlinked by hypertext, and are accessible over the Internet. The resources of the WWW are transferred via the Hypertext Transfer Protocol (HTTP) and may be accessed by users by a software application called a web browser and are published by a software application called a web server.

Appendix 3: Survey

The survey was circulated as an MS Word file with embedded forms. The file is included below.

The Use of Artificial Intelligence in the Audiovisual Sector

Survey carried out for a Briefing Paper for the
European Parliament's Committee on Culture and Education (CULT)

Background and Context

The Committee on Culture and Education (CULT) of the European Parliament is currently preparing a so-called "own-initiative report" on the "Use of Artificial Intelligence in the Audiovisual Sector". This report is related to the CULT Committee's competence on the EU's audiovisual policy in the context of an ever-growing impact of AI technologies on the audiovisual sector, be it *creation*, *production* or *consumption* of audiovisual content.

The CULT Committee asked me to support the preparation of this report with a short briefing paper. In this briefing paper I will focus on the relevance of AI-based Language Technologies for the multilingual, digital European society. In the main part of the paper, I will concentrate on the actual needs and requirements of the European audiovisual sector, as expressed by corresponding representatives. This is why I'm reaching out to you today.

This survey is circulated to a small number of carefully selected practitioners from the audiovisual sector (including media). I'm reaching out to you specifically because we've collaborated in previous projects or initiatives or because you're involved in one of the main EU projects, initiatives or associations in the European audiovisual and media sector, according to my desk research.

Please note that the results of this survey will be reported in a fully anonymous way, which means that no names of organisations or respondents will be reported. Instead, the responses will be aggregated and presented in such a way that the high-level trends, needs and requirements regarding the use of AI in the audiovisual sector become visible. If you decide to participate in this survey, please feel free to shape your responses accordingly.

Thank you in advance for your help, which is much appreciated. Please fill in and return the survey via email by Wednesday, 11 March 2020 eob.[1]

[1] The rather short deadline is due to the ambitious time plan of the CULT Committee for the preparation of the report.

Information about the respondent		
Organisation	Name of your organisation, company, association	
Country	Country in which your organisation is based	
Stakeholder group	☐	Public broadcaster (including television and radio)
	☐	Private broadcaster (including television and radio)
	☐	Film (including wider film production area)
	☐	Video game developer
	☐	Publisher, publishing house
	☐	Telecom provider
	☐	Technology provider, service provider, integrator (including AI tools, Language Service Providers, Analytics, Digital agencies, Multimedia agencies etc.)
	☐	Umbrella association (including community initiatives, community networks, professional associations etc.)
	☐	Research and academia
	☐	Other: Please specify
Name of respondent	Name of respondent (firstname lastname) Please note the remark on page 1 regarding anonymity.	
Role or job title	Role or job title of respondent	

The ten questions of this survey are included in the following pages.

Question 1: Current use of AI		
Question	Do you currently use AI-based technologies in your organisation?	
Answer	☐	Yes (in production use)
	☐	Yes (experimental use)
	☐	No
	☐	Prefer not to answer
If yes	In which areas do you apply AI-based technologies? If you use AI in several different areas, please feel free to focus on those areas that you perceive to be the most important ones.	
Answer	Your answer (maximum of 500 characters)	

Question 2: Future use of AI		
Question	In addition to your answer in Question 1, do you have plans to use AI-based technologies in your organisation in the next 24 months?	
Answer	☐	Yes (production use planned)
	☐	Yes (experimental use planned)
	☐	No
	☐	Prefer not to answer
If yes	Please list the planned areas of application of AI-based technologies. If you plan to use AI in several areas, please feel free to focus on those areas that you perceive to be the most important ones.	
Answer	Your answer (maximum of 500 characters)	

Question 3: Needs, demands and requirements	
Question	What are your organisation's concrete needs, demands and requirements in terms of AI-based technologies (and in which areas)?
Answer	Your answer (maximum of 500 characters)

Question 4: EU policies	
Question	Do you see a need for establishing EU policies (including regulation and recommendations) for the use of AI-based technologies in the audiovisual sector and why?
Answer	Your answer (maximum of 500 characters)

Question 5: Opportunities of AI in the content life-cycle	
Question	AI-based technologies can be used for the (1) creation, (2) production or (3) consumption of audiovisual content. In which of these three broader phases of the content life-cycle do you see the biggest opportunities for AI-based technologies and why?
Answer	Your answer (maximum of 500 characters)

Question 6: Dangers of AI in the content life-cycle	
Question	AI-based technologies can be used for the (1) creation, (2) production or (3) consumption of audiovisual content. In which of these three broader phases of the content life-cycle do you see the biggest danger of AI-based technologies and why?
Answer	Your answer (maximum of 500 characters)

Question 7: European AI tool market	
Question	Please provide a short description of how you assess the market of European AI tool providers for the audiovisual sector (for example, in terms of market access, market size, market diversity, pricing, competitiveness, market dominance etc.).
Answer	Your answer (maximum of 500 characters)

Question 8: Opportunities of AI (in general)	
Question	Independent of your organisation, in which areas do you see the biggest potential or set of opportunities (including technical, social, ethical, economic, legal aspects) for AI-based technologies in the audiovisual sector?
Answer	Your answer (maximum of 500 characters)

Question 9: Dangers of AI (in general)	
Question	In which areas do you see the biggest dangers or threats (including technical, social, ethical, economic, legal aspects) of AI-based technologies in the audiovisual sector?
Answer	Your answer (maximum of 500 characters)

Question 10: European collaboration or coordination	
Question	In which broader areas regarding the use of AI-based technologies in the audiovisual sector do you see a need for increased European collaboration or coordination?
Answer	Your answer (maximum of 500 characters)

This paper contributes towards the identification of current activities, important priority topics and crucial needs in the area of Artificial Intelligence (AI) within the European audiovisual sector. This document was provided by the Policy Department for Structural and Cohesion Policies on the request of the CULT Committee in order to support the drafting of the own-initiative report (INI) "The use of artificial intelligence in the education, culture and audio-visual sectors".

PE 629.221

IP/B/CULT/IC/2020-020

Print ISBN 978-92-846-6695-9 | doi:10.2861/25524 | QA-04-20-257-EN-C
PDF ISBN 978-92-846-6694-2 | doi:10.2861/294829 | QA-04-20-257-EN-N