ARTIFICIAL INTELLIGENCE IN THE MEDIA AND CREATIVE INDUSTRIES

POSITION PAPER

(July 2018)

The position papers has been supported by the Vital Media project, which initiated its creation, contributed to the position paper content, gathered inputs to the paper from the NEM Initiative community, followed up the position paper approval process, coordinated all related activities, and consolidated all version of the position paper.

VITAL MEDIA is a support action project under the Horizon 2020 Programme of the European Union – number: 688310 – project duration: June 2016 – May 2018.

AI in the Media and Creative Industries

Abstract

Thanks to the Big Data revolution and increasing computing capacities, Artificial Intelligence (AI) has made an impressive revival over the past few years and is now omnipresent in both research and industry. The creative sectors have always been early adopters of AI technologies and this continues to be the case. As a matter of fact, recent technological developments keep pushing the boundaries of intelligent systems in creative applications: the critically acclaimed movie “Sunspring”, released in 2016, was entirely written by AI technology, and the first-ever Music Album, called “Hello World”, produced with substantial work coming from AI has been released this year. Simultaneously, the exploratory nature of the creative process is raising important technical challenges for AI such as the ability for AI-powered techniques to be accurate under limited data resources, as opposed to the conventional “Big Data” approach. The purpose of this white paper is to understand future technological advances in AI and their growing impact on creative industries. This paper addresses the following questions: Where does AI operate in creative Industries? What is its operative role? How will AI transform creative industries in the next ten years? This white paper aims to provide a realistic perspective of the scope of AI actions in creative industries, proposes a vision of how this technology could contribute to research and development works in such context, and identifies research and development challenges.

Keywords

Creativity, Artificial Intelligence, Creative Economy, Inclusive Media, Diversity, Personalization, Open Society

Note to contributors: structure of the document and workflow

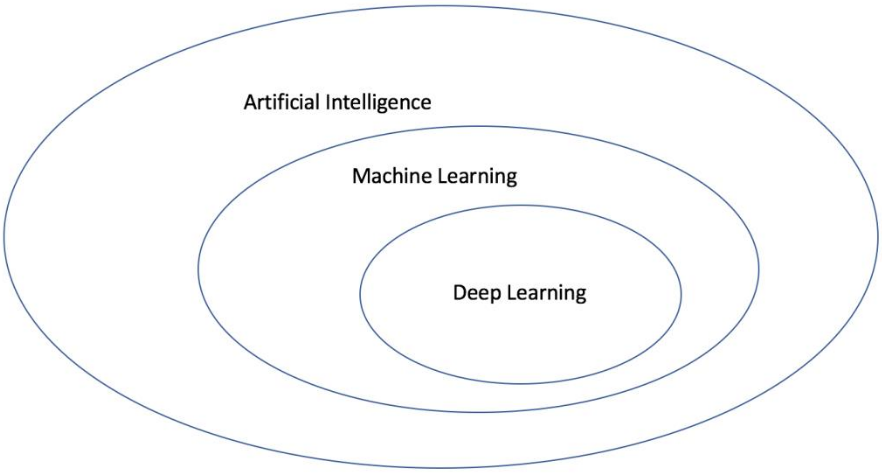
The document proposes to situate AI in the creative industry (Section 2), to describe the most significant use cases organised by application area (Section 3) and to highlight technological challenges (Section 4).

As workflow, we propose to start by completing collegially all the use cases by application area in Section 3. This will provide the paper with an overview of the actual implications of AI in creative industries at large.

Table of Content

1 Summary

Artificial Intelligence is a wide field encompassing several sub-fields, techniques, and algorithms. The field of artificial intelligence is based on the goal of making a machine as smart as a human.



Artificial Intelligence, Machine Learning, and Deep Learning are each a subset of the previous field. Artificial Intelligence is the overarching category for Machine Learning. And Machine Learning is the overarching category for Deep Learning.

Artificial intelligence can be applied to fields as wide as computer audio or visual recognition, self-driving vehicles, robots that can respond autonomously to their environments, recommendations of films via Netflix, and financial analysis.

Artificial Intelligence is usually divided into six categories:

- **Logical Reasoning**. Enable computers to do the types of sophisticated mental tasks that humans are capable of doing.
- **Knowledge representation.** Enable computers to describe objects, people, and languages.
- **Planning and navigation**. Enable a computer to manage mobility from point A to point B.
- **Natural Language Processing**. Enable computers to understand and process language.
- **Perception**. Enable computers to interact with the world through senses.
- **Emergent Intelligence**. That is, Intelligence that is not explicitly programmed, but emerges from the rest of the explicit AI features.

Even with these main goals, this doesn't categorize the specific Artificial Intelligence algorithms and techniques. These are just six of the major algorithms and techniques within Artificial Intelligence:

1) **Machine Learning** is the field of artificial intelligence that gives computers the ability to learn without being explicitly programmed.

2) **Search and Optimization**. Algorithms such as Gradient Descent to iteratively search for local maximums or minimums.

3) **Constraint Satisfaction** is the process of finding a solution to a set of constraints that impose conditions that the variables must satisfy.

4) **Logical Reasoning**. An example of logical reasoning in artificial intelligence is an expert computer system that emulates the decision-making ability of a human expert.

5) **Probabilistic Reasoning** is to combine the capacity of probability theory to handle uncertainty with the capacity of deductive logic to exploit structure of formal argument. The result is a richer and more expressive formalism with a broad range of possible application areas.

6) **Control Theory** is a formal approach to find controllers that have provable properties. This usually involves a system of differential equations that usually describe a physical system like a robot or an aircraft.

Artificial Intelligence, then, appears in some ways to mirror human intelligence. In some ways. And in some circumstances. It usually involves a degree of autonomy and adaptability, and the term is used across a huge number of different computing and non-computing disciplines. It is however, constantly shifting and being redefined, and being applied in a range of different, unexpected circumstances.

2 Situating AI in the creative industries

Artificial intelligence's ability to transform creative working practices has been thrust into the spotlight of late. It is raising through the main activities regarding the value chain of content creation.

2.1 Creation

Arguably, the hardest task for artificial intelligence to take over is content creation. Yet, it is also one of the most crippling challenges for marketers. A number of brands have begun to utilize artificial intelligence in an effort to make content creation quicker and easier. Financial summaries, sports reviews, and other quantitative analyses are ideal for automated narratives generated by artificial intelligence.

For some reason, the fundamental unit of marketing, a piece of content, is still developed as if creativity is a pursuit immune to numbers.

With artificial intelligence marketers can automatically generate content for simple stories such as stock updates and sports reports. AI tools can be put to good use by companies wanting to increase the volume of their content or improve its quality by generating whatever they require.

Advances in Natural Language Processing (NLP), image recognition, and machine learning have given AI the ability to predict what messages and images will drive desired consumer actions. Each engagement with brand content is another data point teaching machines the content they want to see and how they want to consume it.

2.2 Production

Artificial Intelligence (AI) and machine learning could transform the business of producing media. The AI Production seeks to understand how AI could transform the business of producing media. BBC production team is leading the research in this field to identify aspects of their work where they would appreciate help from machine learning techniques, and looking for opportunities to increase the scope and scale of the BBC's coverage by using AI and ML technologies to make media production cheaper and more effective. The results have been good enough to make professionals to be optimistic.

Many production tasks are highly creative, requiring a clear vision, extensive experience and a mature understanding of viewers and listeners in order to craft a programme or package that meets an audience's expectations for a programme while simultaneously captivating them. Equally though, there are many production tasks that are repetitive, or even formulaic. These tasks could instead be performed by machines, freeing up creative people to spend more of their time being creative.

2.3 Diffusion

The primary driver of adoption of AI technology is the opportunity to automate routine workflows that are manually executed. AI also guarantees increasing insights into audiences. These can be used for content monetisation – e.g. in advertising and content licensing – and customer retention. In fact, audience data can be transformed into effective customer retention campaigns or can be fed to personalisation algorithms to establish more personal relationships with viewers, which is key in a direct-to-consumer model.

Content management is a natural area of application for AI technology. Although the unstructured nature of video and audio data makes it more difficult to classify, advances in techniques such as image, emotion and speech recognition have enabled media technology buyers to increasingly rely on AI tools to organise and search their content archives.

Content distribution is another hot area of application with end-users. Algorithms can be deployed to automate and optimize the network efficiency management of a Pay-TV operator, alleviate bandwidth issues in streaming and deliver increasingly personalised experiences to viewers.

The company at the cutting-edge of AI use for distribution is Netflix, which uses AI not only to suggest better content depending on viewing preferences but also to optimise video compression and delivery.

2.4 Consumption

In many interactive systems and medias, AI tools, and notably machine learning, could be used to computationally model the users of such systems. In particular, according to users' input, e.g., their behavior and interactive patterns, a model of the users traits, states, skills and preferences could be built. This model could then be used in order to provide users' with personalized contents and experience, adapted to each user. As an example, for music or movie consumption, a model of the users' preferences in terms of music/movie genre can first be built based on the users' previous choices of music/movie. Then, new music/movie, likely to be suitable to this user's taste could be provided based on recommender systems. For games and education, a model of the users' skills can be built using their past performances at various game difficulty levels (for gaming) or exercises (for education). Then, an optimal sequence of challenges or exercices can be provided to each user, in order to provide the optimal difficulty level to that user, to optimize enjoyment or learning efficiency. Similarly, the users' affective or cognitive states could be modeled according to the users' behaviours and/or physiological signals (e.g., recorded facial expression, heart rate, brain activity) in order to then provide game challenges and training exercises maximizing the user's experience and enjoyment. Overall, AI can be used for modeling the user at two levels: 1) to estimate hidden user states (skills, affective states, cognitive states, etc.) and 2) to learn how to provide optimal content to this user according to these states.

3 A tour of envisioned application areas

3.1 Art

The advent and the development of recording technologies in the XIXth century has undoubtedly created an irreversible revolution in a number of visual and performing arts, by enabling the massive reproduction and wide dissemination of audio, image and video material amidst the general public. Originally based on physical and then analogical electronic devices, the contents thus created are now essentially digital and therefore have become prone to being handled by all sorts of software tools and applications. This enables infinite possibilities in capturing, generating, transforming, combining and broadcasting digital creations at a scale that is unprecedented in human societies.

Whereas, in the XXth century, the typical structure of the visual and performing art industry was based on a well-defined sequential decomposition of roles into creation, production, distribution, and consumption, a strong paradigm shift has begun in the past decade or so, where these demarcations are fading out. Today, more and more music soundtracks are composed and produced within a single framework: a home studio with virtual instruments and mixing tools. Similarly, means of video production have become massively accessible. Distribution is so easy that artists can promote their work themselves through the means of their choice. Even the borders between consumption and creation are getting blurred, as it is becoming more and more common for end users to customize their favorite soundtracks or videos by reordering, rearranging, remixing, or repurposing them in a variety of ways.

In this new context, Artificial Intelligence is emerging as the general framework to support this evolution, as it can provide a wide range of concepts, tools and applications leading to new ways of approaching artistic creation, performance and experience.

3.1.1 Music

Many music-related (and audio-related) fields are currently facing important changes due to the intervention of machine learning and artificial intelligence technology in content processing. The specific challenges of audio content for machine learning relate to handling high temporal resolution and long-term structures. Early advances in machine learning for music were initially borrowed from the field of speech or language processing. Research in the field has recently become more specialized and it has exploded thanks to the creation of massive datasets from music production companies, artist-curated repositories, academic repositories and video streaming platforms. Currently, AI-based technology applied to music has gained interest in a wide range of music-related applications dispatched across creation, production and consumption.

Creation

The typical workflow in computer-assisted music composition is to feed the software program with scores (the input data) of a certain style or by a certain composer. The program extracts composition patterns from these scores and is able to generate new scores respecting these patterns (Briot et al., 2017; Nika et al., 2017). The very same idea is at the core of most of the so-called AI tools in music creation today: a method able to learn the underlying structure in a set of music pieces or sounds, and generates new content that sounds like the music pieces taken as examples. These tools have recently gained in complexity and expressivity, as they spread outside of academia, pushed by new incentives from the tech and music industries as well as the art world[1].

[1] An example is the Magenta group at Google: https://magenta.tensorflow.org/

While the production of the score for a musical track is often one core part of the creation process, a large body of creations are undergone through manipulating audio directly, e.g. when exploiting loops or samples. As a matter of fact, recent advances in machine learning are demonstrating the capacity of modern methods to efficiently process raw audio signals (Van Den Oord et al., 2016), as opposed to MIDI scores only. In this respect, different approaches may be mentioned. First, a large body of research on source separation has recently enabled the *demixing* of music, allowing creatives to reuse only some particular sounds within a track, excluding the rest. Second, generative modeling may be considered to directly produce new musical samples after training on audio datasets.

Creation is fueled with inspiration, for which *style transfer* proved a very interesting technological tool in the domain of image processing, where it enabled new ways of graphical creativity[2]. In the context of music, style transfer would mean transforming an audio piece or a score so that it becomes a representative example of a target style, while retaining its specificities. For instance, transforming rock to tango, saturated to clean vocals, etc. Recent attempts have considered raw audio inputs from the classical repertoire (Mor et al., 2018). Important challenges remain: learning long-term temporal dependencies (whose scope can vary from one style to another), and allowing transfer between very different musical timbres.

In any case, these applications of AI technology to music creation are still at their infancy and can still be considered scientific challenges today. This is first due to the inherent difficulty of generating musical content, which is highly structured and requires high sampling rates, but it is also due to the difficulty of gathering large music datasets on which the systems may be trained, as opposed to the plethora of image datasets available. Certain initiatives already exist such as the AudioSet by Google https://research.google.com/audioset/ that features musical dataset from youtube, but it is far from being an ideal resource for music research, because its core focus in on general-purpose audio processing.

AI-based music creation has also spread outside of academia. The recently released album "Hello World", advertised as the first-ever AI-based music album, involved AI as creativity-support tool, helping an artistic director to generate pieces of sound to be embedded in music soundtrack. In industry, the startup Jukedeck[3] provides musicians with a set of tools able to generate and personalized musical content. The objective is to offer new creative tools to musicians and producers as well as accelerating music making by proposing relevant elements to creatives. Another example is the London-based start-up Mogees[4] that proposes hardware-software solution for musicians to create their own musical instruments by plugging a sensor on everyday objects and by demonstrating to the system how it should sound.

Production

Music production is also experiencing profound changes through the use of AI technology. The current trend for musicians is to work more and more independently from production studios, thanks to the availability of affordable technological tools. A first body of AI-based production systems then typically provide the creatives with audio engineering solutions. As an example Landr[5] is a Canadian start-up that develops solutions for mastering, distributing and communicating new music productions. As for creation, AI-based tools can be promoted as ways for musicians to independently

[2] see e.g. https://www.youtube.com/watch?v=Khuj4ASldmU

[3] Jukedeck https://www.jukedeck.com/

[4] Mogees ltd. https://www.mogees.co.uk/

[5] Landr https://www.landr.com/en

release their music and consequently bypassing the traditional workflow of artistic direction and sound engineering.

The wide diffusion of these large-audience tools are powered by well-engineered API (Application Programming Interfaces) which are a set of functions that can be integrated in third-party softwares and commercialized. An example is the NSynth by the Google Magenta group, able to generate new types of musical sounds and that has been already used in mainstream musical industry[6].

Music production however also remains an industry that requires professional tools. In this respect, rights holders often face the problem of repurposing legacy musical content that has a significant cultural value but a very poor audio quality: many musical standards from the 20th century are noisy, band limited and often only available in mono. There is hence a need for a new generation of tools that are able to enhance such content to make it compliant with modern audio quality standards. AI technology such as audio demixing (see e.g. www.audionamix.com) are promising tools for this purpose, providing professional sound engineers with unprecedented flexibility in audio editing.

Although repurposing legacy content for rights holders is one key application, music creation in the studios or on stage can also strongly benefit from AI technology. In particular, much creativity is lost in the studio when musicians have to record their part independently from one another so as to reduce acoustic interference in the recorded signals. The corresponding recording time is also a waste of time and money for both the artists and the studios. A desirable feature is to process the signals originating simultaneously from all musicians, while preserving audio quality. Similarly, exploiting many low-quality sensors (such as mobile phones) that all take degraded views of an audio scene such as a concert, and combine them to reconstruct a high-quality immersive experience is an important enabling technology.

The core novelty and research challenge in the context of audio engineering is the confluence of AI and signal processing. While signal processing was mostly understood as manipulating audio samples so as to *extract* desired signals from them, AI technology now enables taking signals simply as *inputs* to sophisticated systems that can use training data so as to *extrapolate* information that has been lost and is not present in the input. This line of research is blooming in image processing (see e.g. https://dmitryulyanov.github.io/deep_image_prior) but is yet at its infancy for music processing.

Consumption

Digitization and the Internet already led to a profound change in the way music is consumed, because they enabled the end user to access virtually any music content within a few minutes. Although this was felt as a danger by the music industry for almost two decades, until music streaming services became the core source for funding from the music industry, only recently.

In this context, the added value for selling music moved from providing records in store to providing the users with personalized music recommendations. For this reason, recommender systems became one core activity of companies operating music streaming services such as Spotify, Deezer, Apple, Amazon, etc. Technical approaches for this purpose changed from handcrafted methods to the use of AI technology, exploiting large amounts of user session logs. Music recommender systems are now subject to a blooming research activity.

Another important aspect of music consumption concerns the actual *playback* technology involved. While traditional stereo systems are still omnipresent, a surge of interest in AI headphones or speakers recently appeared, where the loudspeakers are augmented with processing capabilities that enable unprecedented control over the sound. Similarly, it is expected that new software playback

[6] See for instance Sevenim's album created with Nsynth https://sevenism.bandcamp.com/album/red-blues

systems will soon go beyond traditional multiband equalizers to offer more control of the audio stream by the user, which is called *active listening*. For instance, mature demixing technology will soon allow the user to mute vocals from any song in real time, yielding a karaoke version in one click.

From an even higher perspective, we may expect AI to blur the lines between music creation and music consumption, by making it possible for the user to enjoy musical content that has been specifically produced for him/her, based on past choices and user history. With the ability to demix and analyze music tracks automatically also comes the possibility to combine them so as to create new unique tracks. While musicians may produce complete songs as usual, it is likely that artists will shortly only provide some stems, to be used by automatic streaming services to generate automatic accompaniment to the taste of users.

In any case, considering existing musical content as the raw material for future music consumption also opens the path to *heritage repurposing*, where musical archives may be exploited in conjunction with more modern content to yield new and always different musical creations.

As may be envisioned, putting together AI technology and music analysis and synthesis will offer many new perspectives on the way music is consumed, that are totally in line with current trends of adding value through the analysis and the browsing of huge amounts of tracks. The next step forward appears in this sense to also add value through processing and automatic creation.

Blurring boundaries

Creation, consumption and production (See previous section)

References

Briot, J. P., & Pachet, F. (2017). Music Generation by Deep Learning-Challenges and Directions. *arXiv preprint arXiv:1712.04371*.

Mor, N., Wolf, L., Polyak, A., & Taigman, Y. (2018). A Universal Music Translation Network. *arXiv preprint arXiv:1805.07848*.

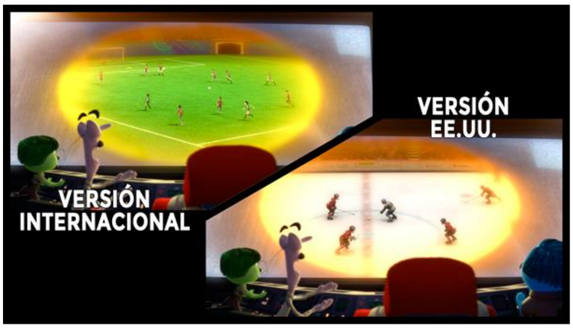
Nika, J., Déguernel, K., Chemla, A., Vincent, E., & Assayag, G. (2017). DYCI2 agents: merging the" free"," reactive", and" scenario-based" music generation paradigms. In *International Computer Music Conference*.

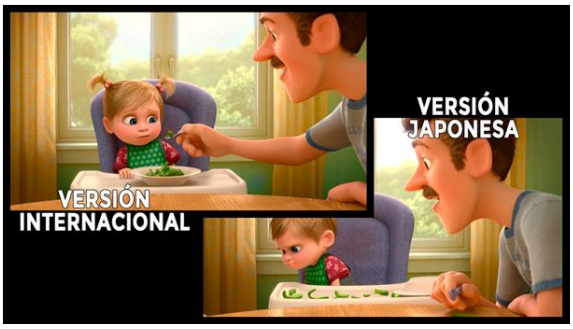
Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.W. & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. In *SSW* (p. 125).

3.1.2 Video, Cinema

Despite cinema and video are a clear target of AI developments, the fight between creators and their natural ecosystem is stopping AI going further in these issues.

Animation is, nowadays, the sector integrating AI naturally.





Depending on language and cultural tradition, part of animation films are converted to adapt to the targeted public, changing idiomatic expressions, sport preferences, gastronomy or gestures.

3.1.3 Images

Creation

AI for generating art images such as photos but also non-photorealistic images is an emerging topic. Leveraging on the impressive results obtained by deep learning methods on production task such as applying for filter and style transfer, approaches have been presented to generate art.

A milestone in the direction of generating art using AI has been DeepDream [Mordvintsev 2015], a Computer Vision program created by Google. DeepDream uses a Convolutional Neural Network to find and enhance patterns in images via algorithmic pareidolia. The input image is substantially modified in order to produce desired activations in a trained deep network resulting in a dream-like hallucinogenic appearance in the deliberately over-processed images. While Deep Dream requires an image as input, the result of the process is so different from the original and so emotional for the viewer to be considered an AI generated art image.

Originally, DeepDream was designed to help to understand how neural networks work, what each layer has learned, and how these networks carry out classification tasks. In particular, instead of exactly prescribing which feature to amplify, they tested letting the network make that decision. In this case, given an arbitrary image or photo, a layer is picked and they ask the network to enhance whatever is detected. Each layer of the network deals with features at a different level of abstraction, so the complexity of generated features depends on which layer we choose to enhance. Thus, the generated image show in each part of the images what has been seen by the network in this specific part even if what has been seen is not likely to be there. Like seeing objects in clouds, the network shows what it sees even if what has been recognized is very unlikely. The dreams that can be drawn by the network are the results of the network experience. Thus, neural networks exposed during training to different images would draw different dreams even using the same image as input.

The interest of researchers on AI applied to art images has recently exploded started from 2016 when the paper “Image Style Transfer Using Convolutional Neural Networks” was presented at ECCV 2016. The proposed method used feature representations to transfer image style between arbitrary images. This paper led to a flurry of excitement and new applications, including the popular Prisma[7], Artisto[8], and Algorithmia[9]. Google has also worked on applying multiple styles to the same image.

Almost all existing generative approaches are based on Generative Adversarial Networks (GANs) [Goodfellow 2014]. The model consists of a generator that generates samples using a uniform distribution and a discriminator that discriminates between real and generated images. Originally proposed to generate images of a specific class (a specific number, person or type of object) between the ones the model has been exposed during training, GAN is now used for many other applications.

Leveraging these results, AI has also been used for generating pastiches, i.e., works of art that imitate the style of another one. In [Elgammal 2017], By building off of the GAN model, the authors built a deep-net that is capable of not only learning a distribution of the style and content components of many different pieces of art but was also able to novelly combine these components to create new pieces of art.

An interesting emerging topic is generating images from captions. Starting from the work [Mansimov 2016], various approaches have been proposed to generate images starting from captions. The goal is to generate photorealistic images, but we expect similar approaches to be applied for generating art images. The generation of high-resolution images is difficult. The higher resolution makes it easier to tell the AI generated images apart from human-generated images. However, recent works by nVidia [Karras 2018] showed exception results growing both the generator and discriminator progressively.

Generating anime faces is the objective of [Jin 2017]. A DRAGAN-based SRResNet-like GAN model was proposed for automatic character generation to inspire experts to create new characters, and also can contribute to reducing the cost of drawing animation.

Production

Production can be seen as the process of creating something capitalizing on something that already exists. In the image scenario, production has various interpretations. It can be seen as the process of editing an image to produce a new one, for instance by using filters or by modifying its content. It can also be seen as the process of using existing images to produce other media, for instance using an image in a video reportage, or in a news.

Artificial intelligence has been extensively used in the image scenario with very significant results, in various applications ranging from enhancing image quality, to editing images, from image retrieval to image annotation. Most of these applications are significant for the production of images.

Artificial Intelligence was successfully applied to reproduce scene-dependent image transformations for which no reference implementation is available, as for instance photography edits of human

[7] https://prisma-ai.com/

[8] https://artisto.my.com/

[9] https://demos.algorithmia.com/deep-style/

retouchers. For instance, in [Gharbi 2017] an approach was proposed that learns to apply image transformations from a large database of input/output image transformation examples. The network is then able to reproduce these transformations, even when the formal definition of these transformation does not exist or it is not available.

Approaches were also proposed to automatically apply photo retouching operators to enhance image quality. The use of photo retouching allows photographers to significantly enhance the quality of images. However, this process is time-consuming and require advanced skills. Automatic algorithms based on artificial intelligence are able to mimic the expert's skill and to provide users with image retouching easily. In [Yan 2014] an approach that combines deep learning and hand-crafted features is proposed to perform automatic photo adjustment. In contrast to other existing approaches, this approach takes into account image content semantics, which is automatically inferred and performs adjustments that depend on the image semantics itself.

Still on the image editing side, recently deep learning based techniques were proposed that allow giving an existing image an chosen artistic style while preserving its content. For instance, it is possible to modify a picture so that it looks like a Miro painting. A significant work in this direction is given by [Gatys 2015]. Here a Deep Neural Network was proposed that creates artistic images of high perceptual quality. The system was trained to be able to separate content and style information in an image, being able to manipulate and produce artistic styles out of existing images.

Artificial intelligence was also used to produce techniques for image inpainting. Image inpainting has the objective to automatically reconstruct missing or damaged parts of an image. Applications examples are restorations of damaged painting, reconstruction of an image after deletion of objects or subjects. In all these case the aim is to modify the original image restoring or editing it so that the modifications cannot be perceived. A relevant approach in this context was proposed in [Yeh 2016]. In this paper, a Deep convolutional Generative Adversarial Network was proposed that is able to predict semantic information in the missing part and to automatically replace it with meaningful content. For instance, if an eye is missing from the image the neural network is able to correctly generate it and correctly place it.

Similarly, also AI-based techniques for image resolution enhancements were proposed. The capability of neural networks to infer semantics in an image was exploited in this scenario to accurately increase the resolution of an image with an excellent quality. In [Ledig 2016], the authors proposed a generative adversarial network also able to recover photo-realistic textures from heavily downsampled images. The proposed approach is able to infer photo-realistic natural images for 4X upscaling factor.

As we stated before, the use of existing images is often necessary for the production of other new contents. For instance, images are often used in reportages, on during the production of news. In these cases, in addition to tools for editing images, also tools to be able to identify and retrieve images relevant to the producer's needs, out of possible very large image repositories, are necessary. Also, in this case, Artificial intelligence has given a significant contribution. Image retrieval is generally performed using images as queries and searching for other images similar to the queries, or using text queries describing the wanted image content. In the first case, which is typically referred as Content-Based Image Retrieval (CBIR), we need a way to compare the query image with the images in the database and to decide which are the most relevant. In the second case, we either need to associate images with textual descriptions, or to generate visual features to be used to compare images, directly from text queries.

For several years CBIR was performed relying on hand-crafted features, that is human-designed mathematical descriptions of image content that can be compared by similarity to judge the relevance of image results to the query image. Recently, a significant step forward was obtained by training deep neural networks to extract visual descriptors from images (Deep Features), encoding significant semantic information. In this case, the high similarity between features is an indication of high semantic relationships between images. Deep features can be extracted using Deep Convolutional Neural Networks, trained to perform some recognition tasks, and using the activation of neurons in an internal layer of the network as features. This is, for instance the approach proposed in [Razavian 2014], where the authors show that performance superior to other state-of-the-art systems was obtained, with the use of deep features.

In order to use text queries to retrieve images, either textual descriptions should be associated with images, or techniques able to generate visual features from text queries are needed. In both cases, artificial intelligence has recently provided significant solutions to this problem.

Artificial Intelligence can be used to automatically analyze the content of images in order to generate annotations [Amato 2017], produce captions [Mao 2014], identify objects [Redmon 2016], recognize faces [Cao 2017], recognize relationships [Santoro 2017]. This information once extracted can be associated with images and used to serve queries.

On the other side, cross-media searching techniques are able to translate query expressed in one media to queries for another media, relying on artificial intelligence techniques. For instance, it is possible to use text to search for images or vice versa. In this case, the advantage is that an image database can be indexed once, using visual features, possibly extracted using deep convolutional neural networks. Improvement of the cross-media techniques, where the vocabulary of terms and phrases that can be translated into visual features is increased, do not require to reindex the entire database of images. Just the query-time processing tools need to be replaced. This direction is pursued in [Carrara 2016], where a neural network was trained to generate a visual representation in terms deep features extracted from the fc6 and fc7 internal layers of ImageNet, starting from a text query.

Diffusion and Consumption

One of the key features needed for an effective and efficient consumption of digital images is the possibility to easily and rapidly identifying and retrieving existing content, which is relevant to one’s needs. However, image content is often not described, annotated, or indexed at the required level of granularity and quality to allow quick and effective retrieval of the needed pieces of information. It is still a problem, for creative industry professionals, to easily retrieve where, for instance, a specific person is handling a specific object, in a specific place. This is generally due to the fact that metadata and descriptions, associated with digital content, do not have the required level of granularity and accuracy.

Professionals that need to retrieve, consume or reuse images, for instance, journalists, publishers, advertisers, often have to rely only on experienced archivists, with a deep knowledge of the archival content they hold, to find material of their interest. However, the amount of material generated and distributed every day makes it impossible to handle it effectively and to allow professionals to easily select and reuse the most suitable material for their needs. This happens because annotating manually, with the required level of detail, the huge volumes images produced nowadays, is extremely time-consuming and thus almost impossible to afford.

Consider, for example, the news production scenario. Every day there is an army of photographers, and journalists, around the world, that send their material to news agencies, related to some event they have witnessed, hoping that their material will be used in tv news, online magazines or newspapers. However, just a small percentage of this produced digital content will be actually published and, often, most of this audiovisual material remains buried in the news archives, unexploited because not easily discoverable.

In this respect, artificial intelligence offers effective tools to address this problem. AI-based tools for content-based image retrieval, for image annotation, image captioning, face recognition, and cross-media retrieval are nowadays available that allow effective and efficient retrieval of images according to user's needs.

Recently, deep learning techniques, as for as for instance, those based on Convolutional Neural Networks (CNN) become the state-of-the-art approach for many computer vision tasks such as image classification [Krizhevsky 2012], image retrieval and object recognition [Donahue 2013]. Convolutional Neural Networks leverage on the computing power provided by GPU architectures, to be able to learn from huge training sets. A limitation of this approach is that many large-scale training sets are built for academic purposes (for instance the ImageNet dataset), and cannot be effectively used for real-life applications.

Face recognition algorithms also benefit from the introduction of deep learning approaches. Among these, DeepFace [Taigman 2014], a deep CNN trained to classify faces using a dataset of 4 million facial images belonging to more than 4000 unique identities. More recently the VGGFace 2 dataset [Cao 2017] was released which contains 2.31 million images of 9131 subjects. A ResNet-50 Convolutional Neural Network was trained on this dataset, which is also able to determine the pose and age of persons.

Challenges

- Generating images from the description is still challenging even if recent works have significantly improved the state-of-the-art.
- In the last few years, the main focus of images generation using AI has been photorealistic images. While the generation of art images have been proved to be possible and relevant, it is still challenging. Instead, style transfer between images can be considered solved and only minor improvements are expected.
- Many production and consumption techniques rely on the capability of automatically understanding the content of the image. AI has significantly increased the type of objects, relationships, actions, events, etc that can be recognized, but the overall task of understanding is still challenging.
- Cross-media search, as using text queries for retrieving annotated images, is a recent promising approach to allow the retrieve of images, out of huge image databases. However, cross-media search is still challenging, especially when issues of expressiveness and scalability are also taken into account.
- Techniques for automatic reconstruction of missing or damaged parts of images work well on simple scenarios as, for instance, faces. Still, improvement is required to work satisfactorily on generic natural scenes.

References

[Goodfellow 2014] J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14)

[Mansimov 2016] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba & Ruslan Salakhutdinov. Generating Images from Captions with Attention. ICLR2016

[Mordvintsev 2015] Mordvintsev, Alexander, Christopher Olah, and Mike Tyka, Inceptionism: Going deeper into neural networks. Google Research Blog

[Gatys 2016] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. Image Style Transfer Using Convolutional Neural Networks. ECCV 2016

[Elgammal 2017] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, Marian Mazzone. CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms

[Jin 2017] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian, Huachun Zhu, Zhihao Fang. Towards the Automatic Anime Characters Creation with Generative Adversarial Networks.

[Karras 2018] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR 2018

[Gharbi 2017] Michaël Gharbi, Jiawen Chen, Jonathan T. Barron, Samuel W. Hasinoff, Frédo Durand, Deep Bilateral Learning for Real-Time Image Enhancement, ArXiv:1707.02880v2

[Yan 2014] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, Yizhou Yu, Automatic Photo Adjustment Using Deep Neural Networks, arXiv:1412.7725v2

[Gatys 2015] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, A Neural Algorithm of Artistic Style, arXiv:1508.06576

[Yeh 2016] Raymond A. Yeh, Chen Chen, Teck Yian Lim, Alexander G. Schwing, Mark Hasegawa-Johnson, Minh N. Do, Semantic Image Inpainting with Deep Generative Models, arXiv:1607.07539

[Ledig 2016] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, arXiv:1609.04802

[Razavian 2014] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, Stefan Carlsson, CNN Features off-the-shelf: an Astounding Baseline for Recognition, arXiv:1403.6382

[Amato 2017] Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, and Fausto Rabitti. 2017. Searching and annotating 100M Images with YFCC100M-HNfc6 and MI-File. In Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing (CBMI '17). ACM, New York, NY, USA, Article 26, 4 pages. DOI: https://doi.org/10.1145/3095713.3095740

[Mao 2014] Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang, Alan Yuille, Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN), arXiv:1412.6632

[Redmon 2016] Joseph Redmon, Ali Farhadi, YOLO9000: Better, Faster, Stronger, arXiv:1612.08242

[Cao 2017] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, Andrew Zisserman, VGGFace2: A dataset for recognising faces across pose and age, arXiv:1710.08092v2

[Santoro 2017] Adam Santoro, David Raposo, David G.T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, Timothy Lillicrap, A simple neural network module for relational reasoning, arXiv:1706.01427

[Carrara 2016] Fabio Carrara, Andrea Esuli, Tiziano Fagni, Fabrizio Falchi, Alejandro Moreo Fernández, Picture It In Your Mind: Generating High Level Visual Representations From Textual Descriptions, arXiv:1606.07287

[Krizhevsky 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105.

[Donahue 2013] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell, DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, arXiv:1310.1531

[Taigman 2014] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14). IEEE Computer Society, Washington, DC, USA, 1701-1708. DOI: https://doi.org/10.1109/CVPR.2014.220

3.2 Interactive Virtual Environments

3.2.1 Games

Consumption

Video games have become one of the main forms of entertainment and a major player in the creative industries. It is currently a market gathering more revenues than other major entertainment medias such as movies and music (web source). The video game sector has been using AI tools for decades now, in particular for designing artificial characters or opponents. However, the recent progresses in machine learning tools and associated data availability opened up the door to more personalized video game experiences. In particular, video games notably aim at providing enjoyment, pleasure and an overall positive user experience to its players (which can also contribute to more sales and/or more subscriptions to the game).

A key element that has been identified to favor enjoyment and a good gaming experience is to favor the state of Flow in the players (Chen 2007). The Flow is a psychological state of intense focus and immersion in a task (any task), during which people lose the sense of time, perform at the best of their capacity and derive the most enjoyment. Flow can notably occur when the challenges offered to each player match their skills, so that the game is neither too easy - which would be boring - nor too hard - which would be frustrating. However, various players have various skills and seek or need different things in a game to enjoy it. Thus, to ensure maximally enjoyable games for a maximum number of players requires personalized games, whose content and challenges adapt dynamically and automatically to each player.

AI methods could be used to design such personalized games, by 1) modelling the users, 2) providing adaptive content dynamically, based on the model of each player (Cowley 2016a), thus favoring flow and game enjoyment. We detail these points below.

First, in order for a game to adapt to each player, it will have to be able to understand this player, and thus to model it automatically from available data. In particular, a useful player model would have information about the player's skills, cognitive states (e.g., attention level), conative state (i.e., motivation) and affective states (e.g., frustration or joy), and possibly the player's traits (e.g., personality). Such states, skills and traits could possibly be inferred from various data available during play, using machine learning techniques (and in particular classification algorithms). For instance, some player's states, skills or traits could be inferred from the behaviour of the player in the game, depending on his/her actions and the context. The skills can also be inferred from the performance of the user in the game, see, .e.g., (Herbrich 2007) and (Bishop 2013) for a simple skill estimator based on machine learning used for Xbox live. Various cognitive, affective and conative states could also be estimated from physiological signals measured from the player, using sensors embedded in gamepads, in gaming devices, or using wearables. For instance, there are first laboratory results suggesting that some cognitive states such as workload, or some affective states (e.g., positive or negative emotions) could be inferred, to some extents, from rather common physiological signals such as from speech recordings, eye tracking, facial expressions (from photos/videos) (Nacke 2008) or from more advanced ones such as muscle activity (electromyography - EMG), heart rate, galvanic skin response (GSR) (Cowley 2016, Drachen 2010, Nacke 2013) or even from brain signals such as electroencephalography (EEG) (Frey 2014, Frey 2016, Mühl 2014). So far, only a few states can be estimated from these measures, and rather unreliably and unspecifically. Modern AI/Machine learning tools hold promises to estimate a larger variety of states (including, for instance, the Flow state) in a robust way, in order to obtain rich and reliable player models.

Then, once these various states, skills and traits identified and estimated, AI could be used to model and estimate how game events, mechanics and/or difficulty levels would impact the player's enjoyment (as reflected by the player's estimated affective states). This would provide a comprehensive model able to predict how the player will behave and perceive the game depending on its states and skills, and depending on the game context. Once such a robust model of the player obtained, different AI tools could be used to provide adaptive content in the game, dynamically, depending on the player's model. Such AI tools could for instance adapt the game difficulty, change or alter the game story or scenario, trigger various events to induce various emotions, or provide new choices to the players. For instance, the player models defined above could be used to predict the possible impacts of various actions and game adaptations, in order to select suitable actions to maximize Flow and game enjoyment. Some authors have recently proposed that machine learning tools such as recommender systems could be used to select actions that have a positive impact on the gaming experience, based on the impact similar actions had on similar users in the past (Tondello 2017).

Extending such promising R&D directions would include studying the use of similar techniques to design serious games, e.g., for education. As it happens, in the field of Intelligent Tutoring Systems (ITS - see Section "3.5 Education"), similar user modelling and adaptive AI tools are used for providing personalized education. Such works also raise some crucial ethical questions. In particular, such player modelling and game adaptation aim at maintaining the player in the Flow zone and at maximizing game enjoyment. As such, there is naturally a risk to lead to game addiction, which was recently recognized as a disease by the World Health Organization (http://www.who.int/features/qa/gaming-disorder/en/). Such AI tools for gaming should thus also include "ethics by design", to also prevent gaming addictions.

References

Cowley, B. U., & Charles, D. (2016a). Adaptive Artificial Intelligence in Games: Issues, Requirements, and a Solution through Behavlets-based General Player Modelling. *arXiv preprint arXiv:1607.05028*.

Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., ... & Ravaja, N. (2016b). The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction. *Foundations and Trends® in Human–Computer Interaction*, *9*(3-4), 151-308.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31-34.

Bishop, C. M. (2013). Model-based machine learning. *Phil. Trans. R. Soc. A*, *371*(1984), 20120222.

Herbrich, R., Minka, T., & Graepel, T. (2007). TrueSkill™: a Bayesian skill rating system. In *Advances in neural information processing systems* (pp. 569-576).

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2014). Review of the use of electroencephalography as an evaluation method for human-computer interaction. Proc. PhyCS, 2014

Frey, J., Daniel, M., Castet, J., Hachet, M., & Lotte, F. (2016, May). Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2283-2294). ACM.

Mühl, C., Allison, B., Nijholt, A., & Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces*, *1*(2), 66-84.

Tondello, G. F., Orji, R., & Nacke, L. E. (2017, July). Recommender systems for personalized gamification. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 425-430). ACM.

Nacke, L., & Lindley, C. A. (2008, November). Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (pp. 81-88). ACM.

Drachen, A., Nacke, L. E., Yannakakis, G., & Pedersen, A. L. (2010, July). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games* (pp. 49-54). ACM.

Nacke, L. E. (2013). An introduction to physiological player metrics for evaluating games. In *Game Analytics* (pp. 585-619). Springer, London.

3.3 Content synthesis for games, movies, engineering and design

Several industries are faced with the challenge of creating large and extremely detailed numerical models of shapes and environments. This considers both the geometry of the objects as well as their surface appearance (roughness, texture, color). For instance, in the video game industry large detailed imaginary places are created, that can be freely explored by users. The movie industry faces similar challenges when virtual sets have to be designed around actors -- this is true both for animated movies, but also for motion pictures in which special effects are now ubiquitous. These environments have to follow precise requirements to produce the desired effects, in terms of aesthetics, navigability, and plausibility (immersion). In design and engineering, when modeling a part, the creative process combines goals driven by structural requirements, functional efficiency, fabrication constraints as well as aesthetics. Technologies such as additive manufacturing allow to manufacture parts with details from a few tens of microns to half a meter.

In all these fields, it has become extremely challenging to produce content using standard modeling tools (CAD/CAM software), even for the most experienced designers. First, the sheer size and level of details requirements make the task daunting: virtual environments in games and movies go from buildings to entire planets. In engineering, finding the right balance between contradictory objectives often requires to go through a tedious trial and error process, exploring for possible designs. Second, the variety of constraints to consider (navigability, structural plausibility, mechanical and structural behavior, etc.) makes standard modeling extremely difficult : the designer often has to imagine what the final behavior may be, try out her designs, and iterate to refine, either through expensive numerical simulation or by actually fabricating prototypes.

To tackle these challenges, the field of Computer Graphics has developed a rich body of work around the concept of *content synthesis*. These methods attempt to automate part of the content creation process, helping the designer in various ways: automatically filling entire regions with geometry, textures or objects, automatically generating detailed landscapes, cities and plants, and even filling building floor prints and generating environment layouts.

AI methods play a major role in simplifying the content creation process. In particular, content synthesis often results in ill-posed problems or complex, contradictory optimization objectives. In addition, finding a unique, optimal (in some sense) solution is rarely the goal. The objective is rather to produce a diversity of solutions from which the user can choose from. This latter point is particularly important in the fields of engineering and design, where *generative design* is increasingly popular: the algorithm cooperates with the user and produces a large variety of valid solutions (in the technical sense) while the user explores and suggest aesthetics. In the video game industry, algorithms that generate playable levels on-demand increase replayability while reducing costs, both in terms of content creation time and in terms of storage and network bandwidth (a game level can take up to several gigabytes of data).

While most technical objectives such as connectivity, structural requirements and geometric constraints can be formulated as objective functions, evaluating aesthetics remains subjective, cultural and personal. In this particular area, AI machine learning techniques are especially suitable. One particular methodology rooted in this trend is by-example content synthesis. These techniques produce new content that *resembles* input data. By matching features such as colors, sizes, curvatures and other geometric properties, the produced content borrows the aesthetics of the input. In this particular area, there is a significant ongoing research effort to exploit latest advances in AI, such as generative adversarial networks (GANs). Clearly, this trend is quickly propagating to all areas of content synthesis.

With the advent of additive manufacturing, this trend is also appearing in mechanical engineering and design. Indeed, one promising way to reduce the weight of parts and achieve novel material properties is to create extremely detailed structures embedded in the volumes of the manufactured objects. This encompasses the concepts of metamaterials, architectured materials and 4D printing, with direct applications in the aerospace industry, medical domain (prosthetics) and automotive industry -- in fact, as engineers are trained in these techniques we can expect the design of *all* manufactured high-end parts to be profoundly revised in light of these possibilities. However, modeling the geometry of such parts is made difficult for very similar reasons: complexity, constraints, details and scale. In addition to these challenges, the parts must now be optimized following complex numerical objectives (structural strength, fluid dynamics, aerodynamics, vibration absorption). This require expensive simulations -- these may be performed on a single object using large computer clusters. However the generative design process requires thousands of these simulations. A key challenge is to exploit AI and machine learning to prune the space of possibilities, and focus full-scale computations only where necessary. Overall, AI will likely play a major role in unlocking the full potential afforded by novel manufacturing processes.

3.4 Societal Challenges for Information and Media

3.4.1 Media Access Services

Multilingualism

Description

The principle of subtitling for the deaf and hard of hearing was introduced in the United Kingdom in the early 1970s to meet the requirements of the hearing impaired people to access TV programs. This first system (Ceefax created in 1972 and Oracle), became widespread on the television channels in 1976.

At that time (1976) appeared in France the Antiope system [1]. Teletext subtitling made its first appearance on Antenne 2 (France Televisions 2 today) on November 1st. 1983, on France 3 and TF1 in 1984, on Canal + in 1994, and on Arte in 1998. The Antiope system has been replaced by the European standard Ceefax on January 1st, 1995.

Today, subtitling for the deaf and hard of hearing is governed by the European standard (EBU / EBU-N19).

Now it becomes necessary to define subtitling in a contextualized way, that is to say taking into account the current state of the art, the technique(s), the regulations and relevant market(s), describing it as:

- A sequence of subtitles that restores the meaning of the speaker's speech while adapting his/her words if necessary,
- Free of spelling errors and misinterpretations,
- Respectful of the standards in the country / countries concerned. These standards are techniques (objective standards), and artistic (subjective rules).

This definition effectively eliminates all automatic captioning providers that abound in the relevant market and do not meet any of the points in the above definition. The future technical challenge is to transform / improve this state of fact.

Challenges

Challenges are about the development of systems for the automatic production of subtitles and sign language of video content using recent developments in Artificial Intelligence applied to machine translation. We have identified 3 major technological challenges:

1. Automatic detection and production of subtitles with regard to multilingual usage: the goal is to develop solutions that automate multilingual production subtitles.
2. Automated sign language representation by avatars: the objective is firstly to produce a sign language representative for video content, and secondly to synthesize this representation through an avatar.
3. Big Data - Real Time: it is important to produce solutions that can handle large volumes of data in acceptable times, and also to build various large corpuses needed to use deep learning.

Fragmented subtitling markets

The market for subtitling is highly fragmented and operates primarily at the national levels. There are no comparative European nor international studies concerning the players involved, their market shares or their intentions for technological development.

According to one of the rare studies on dubbing needs and practices in the audiovisual industry in Europe, there are 631 dubbing and subtitling companies in 31 European countries, 160 of which are leading companies. Their overall turnover was estimated between 372 and 465 million Euros. 84 companies are located in France, Italy, the United Kingdom and Germany, accounting for 64% of the turnover on these activities. 30% of sales (turnover) would be made on audiovisual work.

At the time, as today, the circulation of programs and the transfer of language could be further developed, accessibility is slowly improving in view of the European directives and some dedicated projects that have emerged, but is not applied equally or consistently everywhere. The absence of multilingualism penalizes certain future technological innovation perspectives, and the quality of subtitles is not always present going together with increasing pressure on the translation professionals.

Therefore, future issues are to contribute to:

- Fluidize/streamline the circulation of audiovisual (or video) programs through machine translation, while humans focus on the quality of work, for example.
- Machine translation would also make it easier for television channels to acquire new foreign customers and allow them to invest more easily in extra-European programs without investing too much in subtitling;
- Encourage more synergies and convergence between subtitling and the development of multilingualism or the integration of foreigners (migrants for example) in a given country.

Players involved

Market players are broadcasters (TV, Web, festivals), laboratories, and freelance writers (subtitlers / adapters / translators).

The strong interaction between these three main actors in the economic model of the so-called subtitling service can be presented as follows:

1. A broadcaster orders a subtitling service to a laboratory.

2. The laboratory has the service performed by integrated or freelance authors.

3. The laboratory ensures by a quality control (simulation, correction) that the program meets the standards (it is therefore necessary that the laboratory has clear and recognized internal processes, i. eg. ISO standard).

4. The laboratory delivers a subtitle file to the broadcasters.

Regulation and market opportunity

Legal obligations[2] have been a real opportunity for many actors. In France, the production of subtitles by its specialized service has increased from 6,045 hours at the end of 2008 (made entirely by authors / adapters) to 8,380 hours at the end of 2009 (with 1,939 hours of live programs) and 13 140 hours at the end of 2010 (with 5 097 hours performed live).

Thus, we observe that the legal context has completely shaped and redesigned the production mode (through the introduction of speech recognition software) and the routing of captioning on the air and the production of subtitles. The regulatory issue clearly favors market opportunities.

Finally, the subtitling market will certainly concern cinema and television, but it also concerns more and more advertising and the world of performing arts, education & training, or even the integration of foreigners into a country.

Many countries (France, EU, Australia, Canada, Germany, Hong Kong, India, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Spain, Great Britain, United States, etc.) have adopted legislation similar to French legislation (cited above) on digital accessibility resulting in the need for the production of appropriate subtitles for audiovisual content.

The leading countries in these areas include the United Kingdom and the United States of America. Extrapolating from the French market, we can reasonably estimate that the annual market for the production of subtitles adapted to audiovisual broadcasters at a global level of several hundreds of millions or even billions of Euros per year.

For players in this market, at European level, the Ericsson Group, including its subsidiary Red Bee Media, is a major player in the field of accessibility, providing more than 200,000 hours of subtitling each year, including 80,000 of live subtitling. In the UK, BTI Studios has produced more than 350,000 hours of captioning each year. Finally in the United States, 3PlayMedia is a major player. Among French actors currently known are MFP, ST501, Blue Elements, Dubbing Brothers, Titra Films, Imagine, etc.

Automatic translation out of AI and deep learning instruments will allow to respond to:

• The explosion of content (Big Data aspects);

• Compliance with digital accessibility legislation;

• The reduction of production costs.

IT will redefine this division profoundly because of the emergence of many actors who will constitute as many market targets: universities, administrations, companies, start-ups, producers, etc.

The economic interest is threefold

• First, the automation of the adapted subtitling chain will allow productivity gains that reduce unit costs and increase the volume of processed data.

• Then the production of multilingual subtitles will allow a wider commercialization of the audiovisual contents produced. Distributors of videos and audiovisual programs will be able to market their international programs more easily thanks to the presence of multilingual subtitles.

• Finally, the decrease in subtitle production costs will make captioning accessible to many new players for whom the cost makes captioning impossible.

References

[1] Acquisition Numérique et Télévisualisation d'Images Organisées en Pages d'Ecriture / Digital Acquisition and Televisualisation of Images Organized as written Pages

[2] in France:

https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000809647&categorieLien=id

Inclusion, Diversity, Personalization

Future of personalized access services

Technology is transforming the way we work, live and entertain ourselves. Yet, television (watched on a TV set or via the Internet) is still the preferred medium of Europeans: more than nine out of ten (96%) Europeans watch TV at least once a week. Europeans predominantly watch television on a TV set [1]. But television is changing. It is becoming more connected. Hybrid Broadcast Broadband TV (HbbTV) is an international, open standard for interactive TV, which enables innovative, Interactive services over broadcast and broadband networks [2]. How can the industry guarantee that as many people as possible benefit from this technological innovation? And, if Europe is to become a world leader in accessibility, a topic raised recently by the European technology platform NEM (New European Media, 2016) [3], what steps are still needed?

Between 2013 and 2016, the European HBB4ALL project addressed media accessibility for all citizens in the connected TV/media environment. Its main challenge was to consider the delivery of multi-platform audiovisual content (anytime, anywhere, any device) and make this accessible to all. Access services such as subtitles, Audio Description (AD) and sign language have been available for some decades yet often with little research into how they can be optimized. HbbTV opens up new opportunities for the customization of these services. New access services are also being developed, such as Clean Audio (CA): Following testing as part of the project, the HBB4ALL access services are publicly available on air at RBB and ARD in Germany, at TVC in Spain, at SSR/SGR in Switzerland and RTP in Portugal.

Recommendations for future research and innovations

Accessibility research and innovation issues still to be addressed are:

- Tools to enable increased opportunities for employment in the media and creative industries
- Increased access to digital media services
- Automatic translation to sign language, and from sign language to text
- Automatic translation of subtitles (multi-languages)
- Accessible universal remote control

- Screen reader enabling those with a visual impairment to read subtitles
- Improving multimedia accessibility by design-for-all
- Collaborative work within the industries
- Building on existing media access services and innovation systems (open source and others)

Future short and mid term innovation trends include

- For broadcasting: developing and improving sign language production, Audio Description for content (videos and books) with the facility to deliver dialogue and ambiance elements of the soundtrack separately, achieving robust subtitling performance across genres and increasing interoperability, allowing users to consume personalised automatic live subtitles anywhere.
- For web access developments: industrialize existing prototypes e.g.: subtitle renderer; inlay/screen overlay (incrustation) of sign language interpreter; advanced audio functions; improve the quality of automatically generated subtitles, reliable STT technologies, improve avatar based signing services, develop and integrate additional accessibility services into existing online platforms.

From the user perspectives, it is important to ensure a design-for-all-approach, while recognising that very specific needs may go beyond design-for-all, like affordable, reliable and interoperable solutions, availability of continuous technical support, information about existing and future services, training support for user groups of all ages.

In terms of standards, we recommend to build on European and worldwide standards involving all stakeholders to create large-scale usage. Beyond media accessibility, work on issues surrounding the IoT (Internet of Things), i.e. the interconnectivity of all objects that exchange data, where media access is relevant.

Moreover, clear regulations should exist not only at national level but also at a European level.

It remains crucial to raise awareness in the value chain through information and media, through education and curricula, by bringing stakeholders together (studies, think tanks, projects, market take-up), and include the content production industries.

This shows the need for a continuing emphasis on media accessibility, while recognizing that many strides have been taken in Europe so far. This is best achieved through education; standardisation and legislation based on sound academic and industry research and by the involvement of all members of the value chain, not forgetting the users. If design-for-all is the fundamental principle we will ultimately all benefit from the media interconnectivity. Above thoughts and research recommendations aim to guarantee the future of media access services.

References

[1] Media use in the European Union 2014

http://ec.europa.eu/public_opinion/archives/eb/eb82/eb82_media_en.pdf

[2] https://www.hbbtv.org/overview/#hbbtv-overview

[3] NEM-Access Report: Opening Doors to Universal Access to the Media. February 2016.

http://nem-initiative.org/wp-content/uploads/2016/03/NEM-ACCESS-Policy-suggestions.pdf

3.4.2 News

AI is gradually changing the news media business, impacting all steps from production to consumption.

Creation and production

On the production side, information gathering and synthesis is benefiting and will continue to benefit from increasing technological achievements to facilitate the analysis and cross-examination of heterogeneous information sources in multiple languages, including linked open data and crowdsourcing, to help validating information and facts on a large scale (so-called *fact checking*), to automatically provide insightful, potentially personalized, digests including enlightening visualization and summarization. Examination of the Panama Papers, leveraging natural language processing and text mining techniques in conjunction with database technology and graph visual analytics, is a recent example of this trend[10], which also points at the limitations of today's technology. The recently ended EU project YourDataStories focusing on linked data for investigation journalism is another meaningful example, however limited to homogeneous well-structured data. Addressing technology for heterogeneous sources, the Inria project lab iCODA focuses on the seamless integration and exploration of knowledge bases, public databases and curated content collections for data journalism.

Fact checking is another emblematic use-case where AI is bound to make a difference, as highlighted in recent initiatives such as the EU projects Pheme, REVEAL or InVid. Making use of knowledge representation, natural language processing, information extraction, image retrieval and image forensics deeply modifies the debunking of fake news, while social network analytics provides the means to better understand how and by whom fake news are propagated so as to facilitate their dismantling. On the other hand, image and video manipulation is rapidly improving, in particular with recent advances in deep learning for text and image synthesis (cf. fake discourse of President Obama presented at SIGGRAPH 2017), and fake news producers will sooner or later become aware of the methods used to track them and find workarounds. This calls for the development of efficient countermeasures and adversarial approaches.

Diffusion, consumption

On the consumption side, AI technology also modifies in depth our habits. User profiling and recommender systems are on the verge of being widely used as the number of information sources critically increases. This increase of sources also calls for mechanisms and general public tools for users to assess the reliability of the information they are provided with, beyond the traditional work of press agencies, and possibly across language barriers. News aggregation and summarization is also key in today's news consumption and still requires significant work on automatic multimodal summarization and story-telling easily adaptable to a user's personal expectations, on new content generation, etc. Last but not least, participative journalism is progressively becoming a standard (see, e.g., tweets embedded in newspaper articles or in news shows, videos and photos of events being taken by witnesses and incorporated in professional news reports). This growing trend, which gradually shifts journalist work from professional redaction to the general public, from official news providers to social networks, must be accompanied with intelligent processing tools to maintain high-quality information channels.

[10] http://data.blog.lemonde.fr/2016/04/08/panama-papers-un-defi-technique-pour-le-journalisme-de-donnees in French

Challenges are (to be redacted or moved to the technical section): heterogeneous data integration and querying, with ontology-based access (making the most of participative input, capitalizing on existing knowledge bases and public open data); efficiency, trust and timeliness in information extraction and knowledge discovery (i.e., better collaborative, up-to-date and easy-to-maintain knowledge bases), as well as in content production, whether automatic or not; improve the security of multimedia information retrieval systems and image/video forensics; better personalization and recommendation; trust and transparency of algorithms, and potentially of information (blockchain?) are also at stake here.

3.4.3 Social Media

Social media, today mostly dominated by large companies in the US such as Google, Twitter, Instagram or Facebook, have become an important channel for information and entertainment, conveying huge amounts of personal information that can be used as a proxy to study and monitor people's mind set on a topic or on a product. AI technology has already revolutionized the way social media content is indexed, searched and used, with key technology such as object, face or action recognition in images and videos, entity detection in texts, or opinion mining and characterization. Highly distributed recommender systems exploiting user profiling are today also instrumental to social networks, including for ad placement. Beyond the analysis of user-generated content for indexing and search purposes, monitoring content and users on social networks can provide valuable information and knowledge on specific communities, on people's behaviour and opinions, on societal trends, etc.

Opinion mining and trend detection (elections, marketing, etc.), also comes with community detection. Surveillance: national and EU security but also suicide prevention, cyberbullying detection (e.g., EIT Digital project CREEP). Privacy issues are not to be forgotten. Further info in NEM Social Media Position Paper

3.5 Education

3.5.1 Inclusive education for migrants

Future EU R&D&I funding with personalization and diversity as one key dimension succeed the integration of migrants.

Migration raised major societal challenges within the European Union over the last years, and simultaneously the question of inclusion and access for all through ICT, content and (social) media. Therefore the NEM Steering Board introduced the migration topic, and organized an exploratory meeting between NEM members & UNESCO during the last NEM Summit in Madrid (30 November 2017).

Out of experiences with migrants, it appeared that UNESCO is concerned about the role that technologies may play with regard to integration of migrants into the society. This should be placed under perspectives to promote multiple knowledge (of and about those who arrive), about cultural differences, and the economical sides: many refugees are coming to Europe, and they can impulse economic growth.

In terms of needs, the personalization of services is a first major issue: we are talking about very different profiles of people. Not just one solution expresses and addresses all those different populations. The second issue: mobile technologies can be seen as part of the solution (not THE solution). Almost 80% would do anything in order to have mobile technologies. All refugees care about being connected. So mobile technologies have to be part of the solution.

UNESCO would welcome to get European wide tools to address the refugees' needs. It is about scaling up something that works first, and international cooperation.

Basic and advanced needs are as follows

- Basic needs in terms of education, tools for promoting communication. To help children from Syria to learn Arabic, because they have a lack of education (did not go to school). And also learn the language of the country.
- The remedial needs when it comes to education. Many children will arrive to schools that still recall when their schools were bombed (very traumatic souvenir).
- The higher educational sector. We need systems to help them to certify obtained diplomas. Does anybody address them?
- Consider the opportunity to enhance their digital skills (younger, children, and others). Digital skills are part of to live in the society, and relevant content can help them to be fully citizens in the society.

Tools that UNESCO could support for example are real time translations aiming at empowering human contact. Refugees need to be connected to the country's language, but also to their own language. In addition, the situation requests more qualified teachers (not only those of the receiving countries), but the need of tools and content for (social) media has to be qualified, and extra costs per child (+ 33 to 50 % compared to others) must be considered.

Industries helping to shape solutions would be much appreciated, while UNESCO can welcome any kind of proposal and push it, envisage to help to develop the system(s), and be a partner.

Future R&D1&I related topics could be related on a next public EU R&D&I funding, which considers accessibility, personalization and diversity as one key dimension.

Besides the above mentioned translation tools, media (particularly Public Services) can serve as Educational and Knowledge Diffusion Platform for all - both migrants/immigrants and local population. Content creators, creative people, and storytellers should be "encouraged" to produce content related to immigrants (beyond news and reports based on emergency). Different navigations between different rights and administrative issues need also to be addressed. In the forefront, it is also important to collect research literature to make these topics more predictable.

Higher education is one of the issues: there should be a consolidation of treating people in a non-discriminatory way from both sides. Education institutions have to deal with it from now on. It has to be checked whether there is a lack on tools, or methodology, etc. The language tools here apparently are not the big problem. For children, it might be the EC that comes up with bottom-line instructions.

Concerning innovation and refugees, it is important to:

- Identify ICT instruments for them: at city level, country level, European level. For example security, tracking.
- Innovate with strong PA (Public Authorities) support in order to mitigate possible societal challenges related to "the others" (immigrants).

In the coming month, NEM intends to create a NEM vision taking into account the situation, the needs and further potential NEM solutions, like i.e. create tools to help the programme makers.

4 Identifying technological challenges / Impacts

4.1 AI for humanity

These disruptive technologies indeed raise new ethical and regulation challenges. These aspects are well addressed in several documents, like the Villani et al report on «AI for humanity» and will not be rephrased here. Similarly the General Data Protection Regulation (GDPR) aims primarily to give control to citizens and residents over their personal data and to simplify the regulation. There are however issues that are seldom quoted although essential, especially regarding AI and media.

AI but what for? The 1st question is: What are our real needs ? Are those innovative technologies either answers to real (known or emerging) expectations or offering something new that we consider as a real (may be unexpected) benefit? It is not obvious that our will reduces to adopt all gadgets and reorganize our life as a function of these gimmicks. A critical example is autonomous cars: the global problem on earth is not that cars are not autonomous, but that there are too many, sub-optimally used, impacting too much our environment. More generally, the society we want to build is not a conjuring show, with amazing new tools. There are real huge challenges to take up. One key example, e.g., in education, is to use these new opportunities to challenge education for everyone.

AI can improve media, but media have a major responsibility in contributing to make appropriate collective choices about AI.

AI who is behind? The idea that IA is "automatic" introduces a huge bias, it hides who decides. This goes far beyond the advertising submitted to us. If an IA is retraining my free will, or cheating on me, i can bet that a human is making profit underneath. Moreover, some aspects of the sovereign power tends to be appointed to those how have the power on the data and algorithms, e.g., in education, where the companies building educational digital resources decide in a much deeper way that with books, what is going to be the learning activity. Regarding the media, i.e., this is our freedom of expression that is put in question, and the notion of net neutrality is to be understood at the Internet network level but also at the Web content level as pointed out by the CNNuM in their report oh this subject.

AI can improve media, but media have the duty to make explicit and help us understanding those aspects of AI.

AI … only thanks to GAFAMs ? The fact is Google, Facebook, Amazon and Microsoft GAFAMs are to a certain extends leading the developments in AI, including by sharing open software widely used in research such as Tensorflow or Malmö. Shall we attempt to do better than they do with their huge resources or build on what they share ? The solution might be elsewhere. On one hand, consider GAFAMs want or need to enhance somehow their image with regard to geopolitical issues. On the other hand, clearly state the frontier between what they can propose, and what is definitely the sovereign domain (e.g., education, regulation, or health care). In education the Class´Code project is paradigmatic with respect to Google for Education (GfE) will of leading computer science education in France: this initiative has defended its independence with respect to the “giant”, and proposes a complete French and soon European common good, without refusing to collaborate with GfE, which supports peripheral actions. Because Class´Code gathers more than 70 partners in the related field, it offers an independent leadership on its targeted topic. This could be generalized to other domains.

AI science or belief ? While "weak" IA, (or technical IA, as properly defined by (Ganascia, 2017)) is a reality, and corresponds to the fact that « machines can realize tasks which would have been considered as intelligent if realized by a human » as defined by Minsky, with the intrinsic limit of being always specific to a narrow cognitive task (as formalized by the no-free lunch theorem), what is called "strong" or "global" IA (i.e., that an intelligence with consciousness will (in a near future) emerge from a machine able to reproduce and improve itself) is neither true nor wrong, it is a belief. Very honorable people in the world believe that trees have a soul, and other people believe in strong IA, that will either eliminate the humanity or create a human paradise, for some of us (Ganascia, 2017).

AI can improve media, regarding fake news detection, but media have also to stop propagating fake news about AI.

AI and human though ? As analyzed by (Romero, 2018) in order to cope with these previous challenges, what should we, and our kids, learn IA ? We definitely need to understand how it works, enlightening the fact that massive calculation, making profit of statistics on large data set, provide approximations of cognitive functions. This allows us to realize that this is not magic, and to construct a representation of what can be done or not with it. Regarding computer science in general, we are now beyond the common wrong idea that we do not need to understand how it works but simply (obediently) use it. This also apply to AI principle and it is a major goal for science outreach and science popularization to produce resources for everyone on these topics. Furthermore, in addition to developing computational thinking, this includes developing creativity (and not only using tools as it), developing critical thinking: We must neither be technophobe nor technophile, but technocritic.

AI can improve media, but media should have the goal to help us improving the way we enjoy AI.