AI4media

Empowering Media with Generative AI and Large Language Models: Innovations, Challenges, and Future Directions

White paper - September 2024

Authors: Alexander Plaum - Deutsche Welle (DW)
Angelo Bruccoleri, Lorenzo Canale, Roberto Iacoviello, Alberto Messina, Maurizio Montagnuolo, Fulvio Negro, Stefano Scotta - Rai - Radiotelevisione Italiana (RAI)
Angel Spasov - Imagga (IMG) Rémi Mignot - IRCAM,
Christoffer Holmgård - modl.ai (MODL),
Artur Garcia - Barcelona Supercomputing Center (BSC)

Editor: Danae Tsabouraki - Athens Technology Center (ATC)

Designed by Athens Technology Center (ATC)

Executive Summary

The rapid advancement of Generative AI (GenAI) is transforming the media industry, offering often disruptive opportunities for innovation across various sectors, including content creation, verification, moderation, gaming, and music production.

This white paper explores the multifaceted applications of GenAI, examining both its potential to drastically change media workflows and the significant challenges it presents.

The report provides an overview of how GenAI is being utilised in the media industry today, with a focus on several key areas:

- **Content Verification**: Enhancing the accuracy and efficiency of verification and fact-checking processes.
- **Video Production and Content Automation:** Automating labour-intensive tasks to accelerate production timelines and reduce costs.
- **Content Moderation:** Improving the detection and removal of harmful or inappropriate content on digital platforms.
- **Game Testing and Music Generation:** Streamlining creative processes while maintaining high levels of quality and originality.

The applications of GenAI in video production, content verification and moderation, and music composition can bring tangible benefits to the media industry, enabling faster workflows and cost savings. However, the integration of GenAI is not without challenges. Ethical concerns, particularly related to bias and the risk of AI-generated misinformation, underscore the need for responsible AI development. Technical hurdles, such as maintaining consistency and quality in AI-generated content and managing the high computational demands of GenAI models, further complicate its adoption. Additionally, the operational disruption caused by integrating AI into existing workflows, including issues like vendor lock-in and the necessity for staff retraining, requires careful management. Legal considerations around data privacy, intellectual property, and copyright also present significant challenges that must be navigated to ensure compliance and avoid potential liabilities.

To fully harness the potential of GenAI while mitigating its risks, media industry stakeholders must take proactive steps. This includes investing in the development of ethical AI systems that prioritise transparency, fairness, and accountability, with rigorous testing for bias and the adoption of explainable AI techniques. Collaboration across sectors—between media companies, AI developers, researchers, and policymakers—is essential to advance AI technologies that meet ethical standards and industry needs. Enhancing AI literacy through training programs will equip media professionals with the skills necessary to work effectively alongside AI tools, while establishing robust data governance frameworks will ensure that the data used in AI models is ethically sourced, secure, and compliant with regulations. Furthermore, promoting the establishment of industry-wide standards for AI use, particularly in areas such as content provenance and labelling, will help build trust and ensure the responsible deployment of GenAI in the media industry.

The integration of GenAI into the media industry is both a significant opportunity and a profound challenge. By adopting a responsible and collaborative approach, the media industry can leverage GenAI to drive innovation while upholding high standards of quality, ethics, and creativity. The present white paper aims to offer valuable insights and practical recommendations that guide stakeholders in developing GenAI technologies that not only meet the evolving needs of the industry but also adhere to Europe's ethical, legal, and societal standards.

Contents

Executive Summary 2
Introduction 5
GenAI for Content Verification 6
Status quo 6
Case studies and examples 8
Future directions and opportunities 9
GenAI for High-Quality Video Production & Content Automation 11
GenAI to synthesise content 12
Text to Text 12
Conditional text generation 13
Text and data to images 13
Text to Video 14
Text/image to 3D model 14
GenAI to analyse content 15
Transmodal analysis 15
Tracking topics in textual contents 16
Content watermarking and verification 16
Future directions and opportunities 17
GenAI for Image Recognition in Content Moderation 18
Good data for good AI models 18
Classical image generation techniques 20
GenAI for data enrichment and bias prevention in content moderation 20
GenAI for Testing and Music Generation in Games 24
GenAI for Game Testing and QA 24
State of the art in AI-driven automated game testing 25
Generative AI in game testing 26
Future applications of Generative AI in game testing 26
GenAI for Music Generation in Games 27
Current practices in the creation of music for video games 27
Advantages and limitations of GenAI and LLMs for music creation 28
Current uses of GenAI and LLMs for video game music 29
GenAI for Music Co-Creation 31
Recent evolution of the field (last 5-10 years) 31
Industrial impact: Current models and applications 32
Near term future: Current and future use in real cases 34
Challenges of Integrating GenAI in the Media Industry 35
Conclusions 37
Key takeaways 37
Recommendations 38
Glossary 40
References 42

1 Introduction

The media industry is undergoing yet another transformation, driven this time by rapid advancements in artificial intelligence (AI) technologies.

Among these, Generative AI (GenAI) has emerged as a game-changer, offering new possibilities for content creation, enhancement, verification, moderation, and distribution. From automating complex tasks in video production to enhancing the accuracy of content moderation, GenAI has the potential to transform the way media is produced, consumed, and monetized.

This white paper explores the current state of GenAI in the media industry, examining its applications across various sectors, including journalism, entertainment, gaming, and music. It aims to provide media professionals, AI researchers, developers, and policymakers with a comprehensive understanding of how GenAI is being used today, the challenges it presents, and the future opportunities it holds. The information and insights presented in the white paper come from the extensive experience of the authors in their respective domains of expertise and their work at the forefront of research and development activities within their organisations. Similar matters and strategic research directions for the application of AI and GenAI technologies in the media sector, are also discussed in AI4Media's Strategic Research Agenda.[1]

As with any transformative technology, the integration of GenAI comes with significant challenges. Ethical considerations, such as bias in AI models and the potential for AI-generated misinformation, are paramount. Technical hurdles, including the need for more transparent and explainable AI systems, also pose significant obstacles. Additionally, the operational and legal implications of adopting GenAI in media workflows require careful consideration.

The goal of this white paper is to offer a balanced perspective on these issues, highlighting both the transformative potential of GenAI and the critical challenges that must be addressed to ensure its responsible and effective use. By examining real-world applications and case studies, we aim to provide actionable insights and recommendations for stakeholders across the media landscape, including media professionals, AI researchers and developers, and policy makers. This white paper also seeks to equip readers with an understanding of the pain-points and barriers of using GenAI in different media processes and to offer additional knowledge about the strategies needed to navigate this new frontier, ensuring that GenAI is used not only to innovate but also to uphold the standards of quality, ethics, and creativity in media.

In the following sections, we will delve into the various applications of GenAI in the media industry, explore the challenges and opportunities associated with its adoption, and provide a forward-looking perspective on the future of GenAI applications in the media industry.

2

GenAI for Content Verification

Author: Alexander Plaum - Deutsche Welle (DW)

In this section, we focus on the potential of utilising GenAI and LLMs for content verification – with a reference to DW's, ATC's, and the consortium's own use cases in that domain.

Yasmin Dwiputri & Data Hazards Project / Better Images of AI / Safety Precautions / CC-BY 4.0

We will try and answer questions like: What is the technical status quo in the domain of fact-checking? What are the limitations of AI-driven tools used there? What do GenAI and LLMs bring to the table? What are the benefits and challenges of the new technology? Who is experimenting with what kind of tool? What about the latest trends, potential impact, and future research?

Status quo

Media professionals in 2024 face an ever-growing tide of mis- and disinformation: Manipulated content, fabricated content, content taken out of context – and many other varieties. Whatever the type of false information, it is all getting a boost in scale thanks to AI technology and a wide range of bad actors. As a result, the public is becoming increasingly sceptical of what is presented as "a fact" – and that poses a major problem for democratic societies. Legacy media institutions struggle to keep pace with the real-time nature of online disinformation. Social media platforms and messaging apps accelerate the process of spreading false narratives.

Up until recently, fact-checking and verification organisations mostly relied on a combination of human expertise, digital tools, and semi-automation. More specifically, journalists and investigators would trust their experience and (hopefully) sharp senses while simultaneously utilising software capable of finding sources, highlighting keywords, or analysing patterns – to give just a few examples.

The exhaustive Bellingcat toolkit[2] or the smaller, but entirely self-developed, collection of WeVerify tools[3] (plus respective communities and tutorials) are good examples for this approach. Another one is the suite created by Full Fact[4] and Google, which has also received global attention; it was, for instance, used by Nigerian fact-checkers[5] when their fellow citizens went to the polls in 2023. The claim search, claim alert, and real time transcription functionalities in this example already made extensive use of modern AI technology, but still strongly depended on human investigators setting up and analysing everything. Finally, there is the AI4Media Use Case 1 demonstrator[6], a core outcome of this project. It comprises nine smart verification tools (with a focus on forensics), is even more progressive, and draws on cutting-edge AI models and systems – albeit not the latest generation of them. As an R&D experiment, the AI4Media demonstrator (which also makes use of the established Truly Media platform[7]) has only been operated by test users so far, not by actual investigators in a newsroom.

Unfortunately, in an information sphere influenced by the rise of synthetic media and a global polycrisis, both the "traditional" and the proto-AI concepts outlined above are hitting a brick wall. The sheer volume of information makes it difficult for fact-checkers to keep pace. Standard algorithmic tools struggle with nuanced language, semiotics, and context; they are also susceptible to bias. And more advanced tools, especially in the domain of audiovisual forensics, are always at risk of being outpaced and outsmarted, then failing to detect manipulations and disinformation.

Here is a quick overview of the challenges, as discussed at length in this Reuters Institute article[8]:

- the tools are trained on limited datasets and thus do not perform well on real-world content that includes blurry images and other irregularities
- simple manipulations like cropping or changing the resolution of a file can confuse the tools
- detection results can be misleading – and always require an understanding of how a specific tool works
- deepfake creators can adapt to novel techniques – and make fake content undetectable once again; it seems that the exploitation of generative adversarial networks (GANs)[9], diffusion models[10], and neural radiance fields (NeRFs)[11] were only the beginning.

On top of that, the lines between real and synthetic content are starting to blur: Modern smartphone photo apps tend to use AI for image editing or quality enhancement – which may result in images that are essentially authentic but look somewhat fake. On a meta level, there is the question whether or not it still makes sense to stick to reactive methods, i.e. to the strategy of debunking false narratives after they have already gained traction. Some experts also wonder if the verification game can still be won, considering the (literal) armies of bad actors and the next-level AI technologies enlisted by them.

In any case, fact-checkers and verification experts need to step up their game.

Enter GenAI and LLMs. These technologies promise a new approach to verification and fact-checking. OpenAI's (proprietary) GPT-4[12] or Meta's (open-sourced) Llama 3[13] can be trained on massive datasets of text and code. Despite a couple of flaws (which will be discussed later), they perform rather well when it comes to extracting knowledge, highlighting relevant info, understanding context, compiling and matching entities, answering complex questions, translating, shortening, and generating text. With regard to specific verification tasks, these powerful tools can:

- Extract claims from articles, social media posts, or transcribed interviews.
- Flag claims that seem suspicious based on language patterns or factual inconsistencies.
- Check suspicious claims against credible sources.
- Analyse the credibility of a source by examining historical accuracy, domain expertise, and potential biases.
- Scan and interpret vast amounts of information in a short time (thus also allowing real-time checks).
- Generate clear and concise summaries of fact-checks (thus aiding in the dissemination of truth).

Case studies and examples

One of the earliest fact-checking experiments involving a cutting-edge LLM was carried out by Politifact in early 2023. The organisation explored if ChatGPT was able to do the job of a human fact-checker.[14] The engine performed poorly. It got some answers right, but also made a lot of mistakes due to lack of recent information, inconsistency in responses depending on question phrasing, and a focus on providing users what ChatGPT thought they would like to hear. The system also had difficulty understanding context and nuance, made factual errors, and invented information or so called "AI hallucinations".

In an interview on the Inria Blog[15] published in early 2024, representatives of the French research institute and the broadcaster France Info had more positive news to share regarding AI and fact-checking. They mostly discussed the functionality and solid performance of two tools developed by Inria and partners: StatCheck and ConnectionLens. StatCheck is an automated fact-checking program that verifies information by comparing it with large amounts of data (e.g. official statistics databases). It can also understand natural language (e.g. to analyse social media posts). ConnectionLens is a tool that can be used to cross data sources in all sorts of formats and from all sorts of origins. It is useful when it comes to identifying potential conflicts of interest, for example.

On April 4th, 2024 (International Fact-Checking Day), Indian media organisation Factly[16] introduced their work on the Poynter portal.[17] Factly, which has to address the challenges of misinformation in a country with many languages and a large internet user base, currently uses two main products driven by AI: The first, Dataful, seems quite similar to the software Inria has been building: It is basically a portal that provides access to public government data sets. This makes it easier for fact-checkers to find the information they need. The second product, Tagore, shows similarities to ChatGPT, but there is much more focus and customization. Factly describes the tool as a generative AI platform that uses chatbots. These chatbots are built on custom databases that the organisation has created over time. There are different chatbots that focus on different areas, such as SACH (that goes through fact checks) and Parlens (that searches data from the Indian Parliament).

An interesting special purpose tool harnessing the power of LLMs was first introduced by DW Innovation in late 2023: SPOT is A Natural Language Interface for Geospatial Searches in OSM[18] and helps investigators find and investigate a scene of news – and it does so much faster than a "traditional" tool like Overpass Turbo.[19] The software is still in closed beta, but already works quite well (according to test users). Another promising geolocation tool driven by cutting-edge technology is GeoSpy[20], which uses GenAI to try and "guess" where an image was taken. Experts (e.g. at DW Innovation) say that it is quite good but suffers from the typical GenAI inconsistencies.

The lessons learned in the case studies and examples introduced here can be summarised as follows:

- Generative AI and LLMs can make some parts of fact-checking and verification faster and more efficient
- Generative AI and LLMs currently lack the ability to accurately determine what is true and what is false – and run the risk of producing hallucinations
- Human fact-checkers remain essential for critical thinking, understanding the entire workflow, and ensuring accuracy

Future directions and opportunities

Unsurprisingly, the field of generative AI for verification and fact-checking is evolving fast.

Key trends include:

- Fact-checking with **retrieval augmented generation** (RAG)[21]: Researchers and practitioners have started working on concepts that combine more traditional database queries (which produce reliable answers) with LLM architectures (which guarantee natural language communication and output that is easy to understand).
- Fact-checking with **explainable AI**: Researchers have started looking into methods and features that allow LLMs to explain their reasoning and present some sort of evidence trail. This 2023 paper[22] provides a survey of what is called "rationalisation for explainable NLP".
- **Multilingual disinformation detection**: As false narratives transcend borders, projects like the Multilingual Fact-checker Chatbot[23] (funded by EMIF) try to offer AI-driven fact-checking assistance to a global target group.
- **Advanced deepfake and synthetic media detection**: Several research groups are working on tools that can detect highly sophisticated fake media content. A prominent example is FakeCatcher[24] built by Intel in cooperation with SUNY Binghamton. Among other things, the software analyses pixels indicating the blood flow of a subject to see if the image is real or synthetic.

It is also worth noting that major players like Open AI have started working on detection tools for their own GenAI (e.g. there is the DALL·E Detection classifier[25], which unfortunately is a very mixed bag[26]). At the same time, IT companies like The Hive, who train their own models and offer all sorts of AI-driven services to their customers, have included "AI-Generated Image Detection" as a standard service in their (commercial) toolkit.[27]

As for **alternative approaches** that focus on the verifiable **origin of content**, it is important to mention the Coalition for Content Provenance and Authenticity (C2PA) – which basically combines the work of the Content Authenticity Initiative[28] and Project Origin.[29] In this alliance, big tech and media players like Adobe, Microsoft, or the BBC are trying to create standardised ways of embedding metadata and labelling trusted sources. The general idea is to let users check where an asset comes from, when and by whom it was modified – and if there are any irregularities. At the same time, companies like Liccium[30] work on the ISCC[31] standard which can verify content ownership of photos and videos, even if they no longer contain metadata.

Many journalists and educators also push for more **AI literacy** – and better **media and information literacy** (MIL) in general. They explain how to spot deepfakes (e.g. in these posts by MIT Lab[32], DW[33], or AP[34]), curate entire learning guides (e.g. Tackling Disinformation[35] by DWA) or host workshops for organisations dedicated to critical thinking (e.g. Lie Detectors[36]).

Coming back to the approach of verifying/falsifying via technical content analysis – i.e. the focus of this paper – we can state that the integration of GenAI seems promising. Its **impact** may include:

- A reduced fact-checking workload: automating laborious tasks like claim identification and basic verification can free up human expertise for in-depth analysis.
- An increased fact-checking capacity: AI solutions could empower smaller organisations with limited resources and help fill blank spots on the global verification map.
- An enhanced public trust in information: By providing faster and more transparent

fact-checking, generative AI can contribute to a healthier information ecosystem.

To harness the full potential of generative AI and LLMs in verification and fact-checking, **collaboration between various stakeholders** is crucial:

- Joint initiatives of tech companies and research institutions (potentially funded by the Horizon Europe[37] or EMIF program[38]) need to explore advanced AI architectures and interfaces.
- Fact-checking organisations and media outlets (organised via the IFCN[39] and similar groups) need to ensure AI-powered tools align with real-world needs and journalistic standards.
- Policymakers and regulators need to establish guidelines and laws for responsible AI development and deployment.

In summary, it is safe to say that GenAI and LLMs will soon become essential in complex, potentially global verification workflows. They can provide insights and much needed assistance where more "traditional" solutions fail. As a matter of fact, investigators are likely to be rushed off their feet without the new tools. However, the technology still has inherent flaws, and much more research is needed to ensure a responsible and effective use. There are a lot of ethical, technical, and social challenges to be considered (see Section 7 below, where common challenges are discussed in more detail). Users must stay alert, cautious, and sceptical. Developers must strive for transparency, explainability, and good design. In any case, GenAI and LLMs can only be a success in the domain of verification if academia, industry partners, and media experts collaborate. Developed and deployed in the right way, technology can play a significant role in the fight against disinformation – and for a more informed public.

3

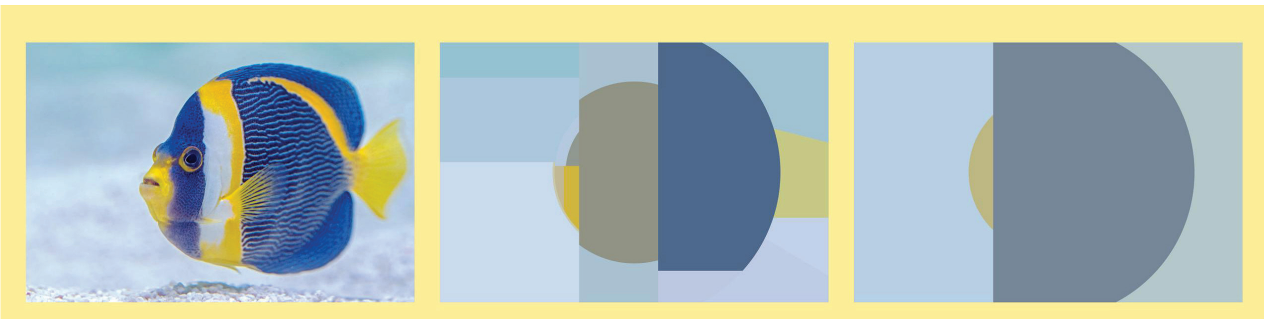
GenAI for High-Quality Video Production & Content Automation

Authors: Angelo Bruccoleri, Lorenzo Canale, Roberto Iacoviello, Alberto Messina, Maurizio Montagnuolo, Fulvio Negro, Stefano Scotta - Rai - Radiotelevisione Italiana (RAI)

As many other business sectors nowadays, the Public Service Media (PSM) is going through a profound transformation ignited by the introduction of Generative AI systems, including image and sound generators, and Large Language Models.[40] Examples of applications include news editing and publication[41], archive exploitation and reuse[42], or audience engagement and marketing analysis.[43]

We believe that the fundamental reason for this is to be sought in the astonishing potential of these technologies to bridge the operational gap between the mere existence of a piece of technology and its efficient employment in industrial processes. With a small amount of code, it is today possible to write complete computer programs solving tasks that were simply impossible to think about just a few months ago. Nevertheless, while some of the usual pains suffered by developers who wanted to introduce information automation in the media processes appear notably diminished, other challenges rise in fields previously lying in the background.

Far from illustrating a complete landscape of the matter, the following pages are rather meant at identifying and briefly describing which operational contexts around the High-Quality Video production & Content Automation use case represent, better than others, environments where GenAI in general - and LLMs in particular - can flourish. A non-exhaustive list of application domains that could benefit from the introduction of GenAI tools is also briefly presented and discussed. Given the scope of this work, the approach followed is not based on a market survey or a technological landscape analysis, but rather on the observation, from the privileged standpoint of a R&D department of a PSM, of potential business benefits, as well as on the objective evaluation of current processes' shortages and inefficiencies.



Rens Dimmendaal & David Clode / Better Images of AI / Fish / CC-BY 4.0

GenAI to synthesise content

The generation of "new" content can be considered as a daily business for a media firm. From news coverage, to documentaries, fictional content, advertisement, and all other kinds of production, regardless of the final publishing media (e.g., broadcast, online, social), creating new audiovisual essence has been and is still the core of the business. To make all this work, literally hundreds of person-years of studies, skill refinement, and on-the-field operation have been spent and a lot of money has been invested. What further or different benefits are we expecting, then, from the integration of a paradigm based on the employment of AI technology in this context? We propose some hints in the following sections.

Text to Text

The introduction in the market, since the middle of 2010's, of online digital services gave birth to a completely new fruition paradigm, based on personalised on-demand access to content. To support such a process, it is necessary that published media are accompanied by rich metadata, which – in turn – are quite expensive to produce.

Since it is rather impossible to foresee in advance any specific usage condition of the published descriptive information, it is an obvious choice to opt for aseptic and quite generic descriptions. This choice, however, does not optimise the utility function value of the descriptive content, or at least not for all possible user categories. An example of employing AI to alleviate this problem could be that of the dynamic adaptation of textual descriptions of online resources for enhanced personalisation and accessibility. Being able to capture essential indications from the user's profile and turn them into actionable instructions to adapt descriptive texts cannot but increase this value and – on average – boost the performance of media services. Figure 1 shows an example of such a service: the transcript of a documentary is firstly summarised and then adapted according to a user's profile.

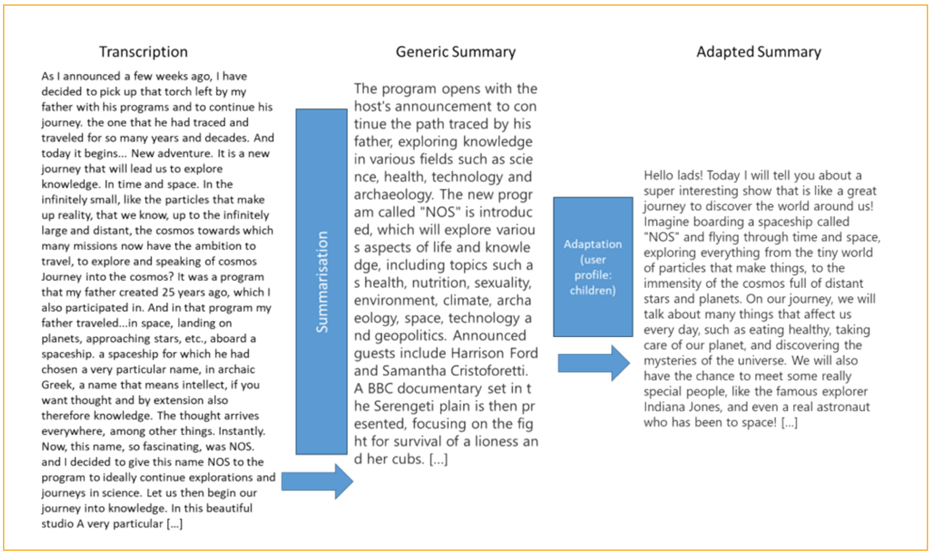


Figure 1 Example of text adaptation

Conditional text generation

Differently from pure text-to-text, conditional text generation is the process by which a language model is used to generate text based on an input instruction provided in the form of a prompt. Out of the tons of possible examples, we could mention the specific case of generating scripts and related acting roles from the authors' indications expressed through simple outlines and/or subsequent chatting.[44] Naturally, a text-to-text variation of such a process can be envisaged when the starting point is an existing script (or extended description from a different manifestation – such as a book text or song lyrics), which is used to produce a new script that adapts the original work to be published on a new media. The above scenario can find realisation in the media business, and particularly in the PSM case, considering the sheer amount of archived material notably preserved by these organisations that can be brought to new life through these means.

Text and data to images

It is commonplace that modern communication has been transformed by the advent of social media and the Internet in a way that is much more important than any previous major technological breakthrough. The amount of new data available online is increasing at a phenomenal pace, which is orders of magnitude higher than the ability of commentators, authors, and journalists to increase their information processing bandwidth.

In the presented context, an example could be that of making use of advanced data-to-image technology to illustrate complex phenomena, abstract arguments, or other high-level concepts with catchy images conveying key semantic information related to those concepts.

A case could be that of constructing illustrative graphs, conceptual maps, and other visual artefacts out of a collection of textual documents which relate to a specific current event or to a wider phenomenon. Substantially based on the Retrieval Augmented Generation (RAG) paradigm[45], this content can have a variety of embodiments, styles, and properties, all sharing a common trail in the ability to capture and deliver novel insights on a complex matter. Figure 2 presents an example of another application in which different GenAI capabilities are used in a chained fashion. We asked GPT4o to write a caption of an iconic image representing the programme summarised in Figure 1, then we used DALL-E to generate the image. Although results generated this way very often seem somewhat naïve, one could use them as an inspirational input to artists in charge of providing contextual visualisation for specific programme content.

Image of a futuristic spaceship called "NOS" flying in space, with planet Earth visible in the background. The spaceship has a modern and attractive design, with bright lights and colorful details. Around the spaceship, there are various elements that represent the different themes of the program:
An enlarged molecule floating near the spaceship, symbolizing the world of particles.

A planet with rings, representing the exploration of the cosmos.

A healthy food icon, such as an apple or carrot, symbolizing the importance of eating healthy.

A small tree or leaf, representing healing for the planet.

A stylized image of Indiana Jones with his classic hat and whip.

A figure of an astronaut wearing a helmet, floating in space.

Figure 2 Example of text-to-image

Text to Video

By analysing existing media and data, GenAI tools are able to create realistic videos from textual inputs. For instance, they enable the production of lifelike weather effects and crowd scenes, offering a degree of realism that's challenging to reach and time-consuming with traditional methods. Moreover, they allow content creators to implement on-the-spot changes to their production, eliminating the need for costly additional shooting.

Another field of application is the creation of animations. This can be done in several ways.[46] One such method is the text-to-animation process, in which users generate an animation by entering a text prompt and adjusting settings. Another approach involves combining text with an image. In this case, users start with an image and blend it with a text prompt to animate their initial idea. A third technique consists in enhancing videos with text. Here, users begin with a video and fine-tune it with text prompts to guide the animation towards the desired outcome.

Despite the impressive capabilities of such tools, there are limitations to what they can achieve, and sometimes they do not meet the exact expectations of filmmakers or content providers. For instance, consistency in text-to-video generation with GenAI is a significant challenge due to the need for spatial and temporal coherence across frames. This requires maintaining long-term dependencies, which can be computationally expensive and difficult to scale.

Text/image to 3D model

Looking at the near future of broadcast production, synthetic media tools are carving out space to support programme makers and broadcasters in enabling and extending 3D content creation and storytelling options including:

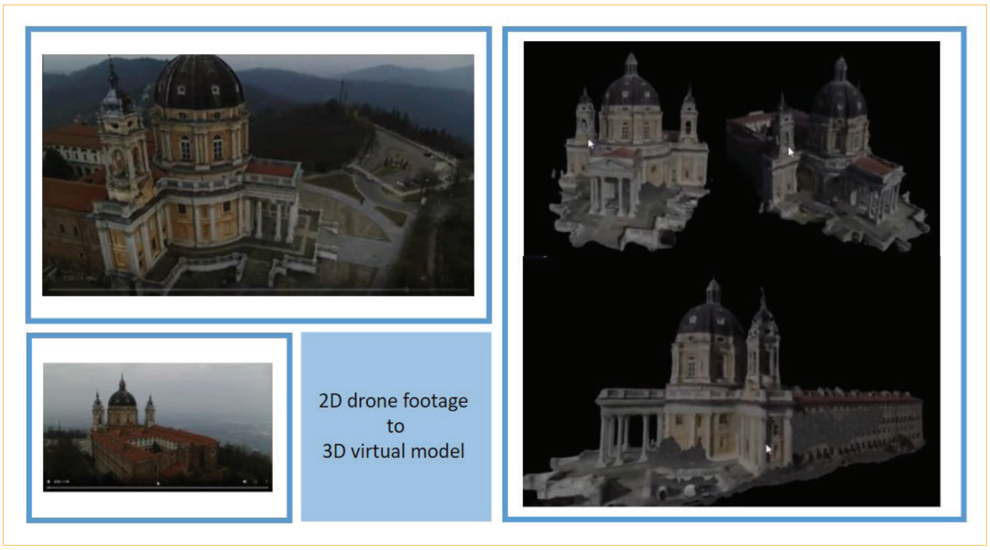


Figure 3 Example of image-to-3D model

- The creation of digital humans, e.g., remote performers or sign language interpreters.
- The creation of both realistic and fantasy environments for human performers, synthetic performers, or a combination of the two.
- The automation of animation productions.

The rapid generation of 3D assets from textual inputs allows users to easily express their ideas without the need of advanced 3D modelling skills, reducing manual work, and speeding up content creation. However, models generated today may lack fine details or realism. This is partly because textual descriptions can be ambiguous, resulting in imperfect 3D translations.

Image-to-3D model tools (an example is shown in Figure 3) capture intricate details from images, preserving lighting, textures, and proportions.[47] However, they require advanced algorithms to extract depth information. The quality of the resulting 3D model depends on the resolution and clarity of the image. Post-processing is often necessary to refine the models.

As far as the generation of digital humans is concerned, as humans we are very sensitive and disturbed by things that are not quite right or uncanny. Scientifically, the point at which realism becomes too realistic, but not realistic enough, creating a lack of affinity, is known as the 'uncanny valley', and was first mentioned by the Japanese robotics professor Masahiro Mori in 1970.

GenAI to analyse content

Similar to creating new content, it is essential for PSM companies to quickly and precisely analyse content already available (e.g., in their archives), which constitutes their real historical memory, as well as one of their main cultural and economic assets. In this context, GenAI tools stand out for their usefulness, providing solutions to analyse content in different ways and from different perspectives. Some examples are reported in the following subsections.

Transmodal analysis

The first aspect to be considered in the content analysis domain is that GenAI enables more comprehensive ways of approaching this analysis. Recent more advanced approaches are **multimodal LLMs**, i.e., models that rely on independent processing pipelines to extract information from multiple channels (e.g., textual, acoustic, visual) and then elaborate the resulting data in a common hybrid token space, normally providing a final textual output. In addition to this approach (performed e.g., through Llava, GPT4o, Gemini), **transmodal analysis** is an alternative that operates by generating textual descriptions conveying contributions from all available channels and making textual LLMs process them towards different objectives (see Figure 4). Recent experimentations conducted in AI4Media showed that such an approach is promising for tasks like the editorial content segmentation, where it is essential to consider contributions from various media channels in the context of a specific analysis purpose.

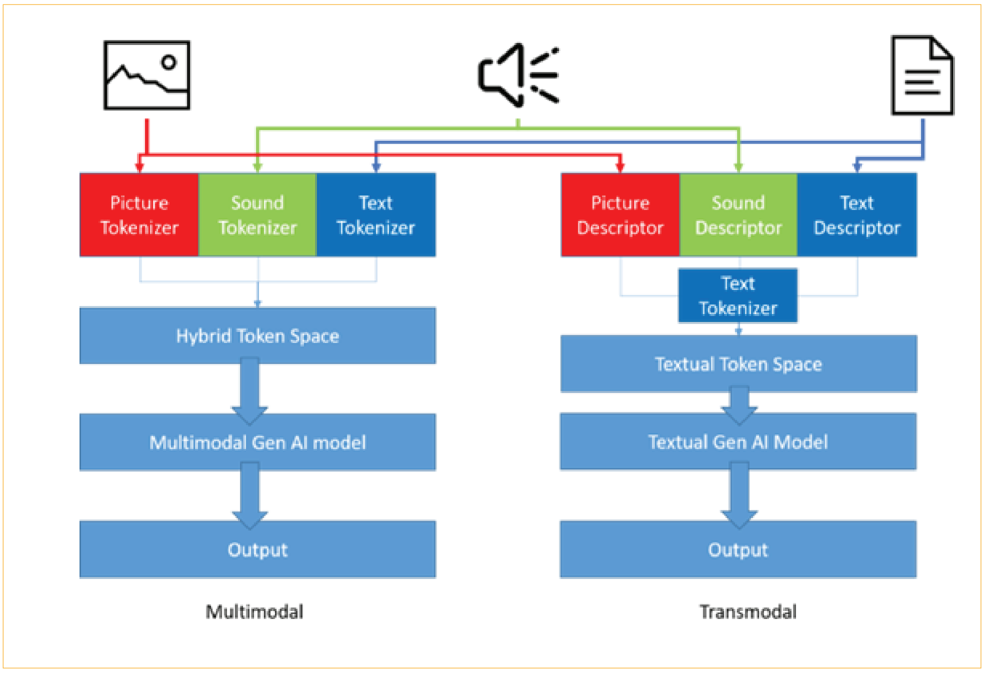


Figure 4 Multimodal Vs Transmodal approach

Tracking topics in textual contents

A useful task in which GenAI and LLMs in particular can help is the semantic analysis of textual content. Indeed, the most powerful LLMs are very good at understanding the main topics of a text, even if these topics are not always explicitly mentioned.[48] PSM companies could use this feature not only to analyse written texts, such as news articles, but the transcripts of media content as well. A possible schema to use LLMs to detect topics treated in a text could be the following (see Figure 5):

- First, split the full text in shorter segments according to the context capacity of the chosen LLM, and to the precision of the required analysis. Having short segments allows users to better know what is analysed in each part of the text. On the other hand, the risk is to lose some background topic whose detection is difficult to catch in shorter segments.
- Afterwards, prompt the LLM to detect topics of interest in each segment.

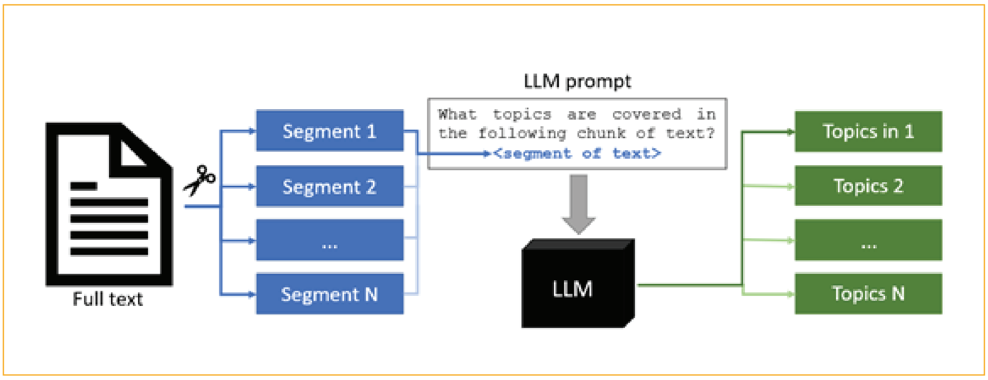


Figure 5 Schema of topic detection within a textual content with LLM. The prompt is simple and too vague to be generic

In the text chunking step, it is essential to reach a balance between topic uniformity, which increases with shorter chunks, and full topic characterization, which increases with larger text portions. Ideally, chunking should be operated at topic borders, and mostly independently from the specific topic (in a non-contextual topic change detection fashion). In the prompting step it is fundamental to define what kind of topics the user wants to detect, feeding the prompt with an as precise as possible definition of what they are looking for.[49] Indeed, simply asking the LLM to return the detected topics could result in an explosion of different (even if slightly) topics that could then be difficult to manage.

Content watermarking and verification

In the sections above, we presented several examples of how GenAI and LLMs can be used to create new content and transform an existing one. There is no doubt that with growing computational capabilities and technological developments, there soon will be no boundaries to the amount and type of content that can be generated. This poses the problem of differentiating human-created content from content generated by AI, as well as of labelling and tracking it.

The fulfilment of these needs will likely require the use of watermarking techniques to ensure data provenance, integrity, and traceability. **AI watermarking** is the process of embedding a digital fingerprint into GenAI and LLM generated content.[50] This fingerprint, which is invisible to humans, is detectable by algorithms, so that the content (and the AI model that generated it) is always recognisable at any point in the content distribution and fruition chain. As a result, policy makers around the world are considering how watermarking techniques should be designed and implemented.[51]

Future directions and opportunities

This brief account illustrated how generative AI technologies can play a crucial role in the future of media business. The number and relevance of possible use cases, reported here only in a small portion of those possibly imaginable, witness how they can represent a true breakthrough in terms of business support and development, both for existing and new services. Challenges, on the other hand, are evident when it comes to the management of these components at various levels. In fact, ethical concerns are of utmost importance and technical issues arising from the integration of these components are far from being easily addressed yet. From the point of view of a digital media company there is no doubt that GenAI will represent a key opportunity (rather than a threat), provided that adequate acknowledge is given to the following aspects:

- Necessity to adopt a usage- and user-driven approach at the adoption of GenAI technology, starting with a clear definition of the functional requirements that such components are supposed to fulfil in each target process.
- Necessity to build and maintain a common culture at the enterprise level, acting as a sort of immunological barrier against misuse and misconception of what generative AI is able or not able to do.
- Maintain a strong commitment on internal staff working on the topic, to avoid vendor lock-in situations and underestimation of the negative impacts that imperfect or unaligned tools may have on the processes due to their inherent operational gaps.

From the point of view of innovation and applied research, we believe that – among many others – a promising future direction would be that of **AI agents and agent communities**, especially those configurations in which different kinds of AI (descriptive, transformative, generative) are integrated in a self-orchestrated workflow. This is mainly due to the observation that media processes are typically a complex mixture of automation and creativity, a distinctive element compared to other businesses. In such an environment, the various kinds of AI tools can offer the right amount of flexibility and autonomous behaviour, best fitting the dynamic and ever-changing requirements of typical modern media workflows.

AI's rapid advancement across various forms presents immense potential for enhancing creativity, driving innovation, and boosting productivity throughout broadcast production. However, this technological progress also comes with notable and unprecedented risks and challenges. On the one hand, any implementation of AI in PSM must align closely with the values and mission of these organisations, as AI technologies are more and more embedded in tools from external vendors or in freely accessible online platforms. Therefore, it's crucial to carefully consider the company's data privacy management rules when using them. Special attention must be given to protecting sensitive information and ensuring compliance with data protection regulations. On the other hand, the integration of GenAI technologies into PSM companies' workflows, with a specific focus on LLMs, poses technical challenges, including cost, performance, and quality considerations.

4

GenAI for Image Recognition in Content Moderation

Author: Angel Spasov - Immaga (IMG)

Anne Fehres and Luke Conroy & AI4Media / Better Images of AI / Data is a Mirror of Us / CC-BY 4.0

The primary goal of Image Recognition models is to accurately detect objects or concepts of interest in visual content.

The performance of these models heavily depends on the quality of the data used for their training, yet obtaining high-quality data is often a challenge. Recent advancements in Generative AI offer new opportunities for Computer Vision engineers to enhance their datasets. This chapter explores the challenges and opportunities presented by this new paradigm, focusing on its application in developing Content Moderation systems.

Good data for good AI models

At the heart of content moderation systems lie AI models. Those targeting visual content must be capable of recognizing defined objects, such as radical symbols, flags, signs, or more general concepts such as human rights violations in images and videos.

But first, the AI models need to be taught what these objects or concepts are and what they look like. To achieve this, the models require images and videos where such concepts are present (i.e., positive examples) and where they are missing (i.e., negative examples). This data lays the foundation of every computer vision model but is often arguably the most difficult part to get right and the most common reason for poor model performance. Essentially, it is no exaggeration to say that an AI model is only as good as the data that has been used for its development. But what does it mean to have good data?

There are several aspects that good data needs to comply with. It is crucial for the data

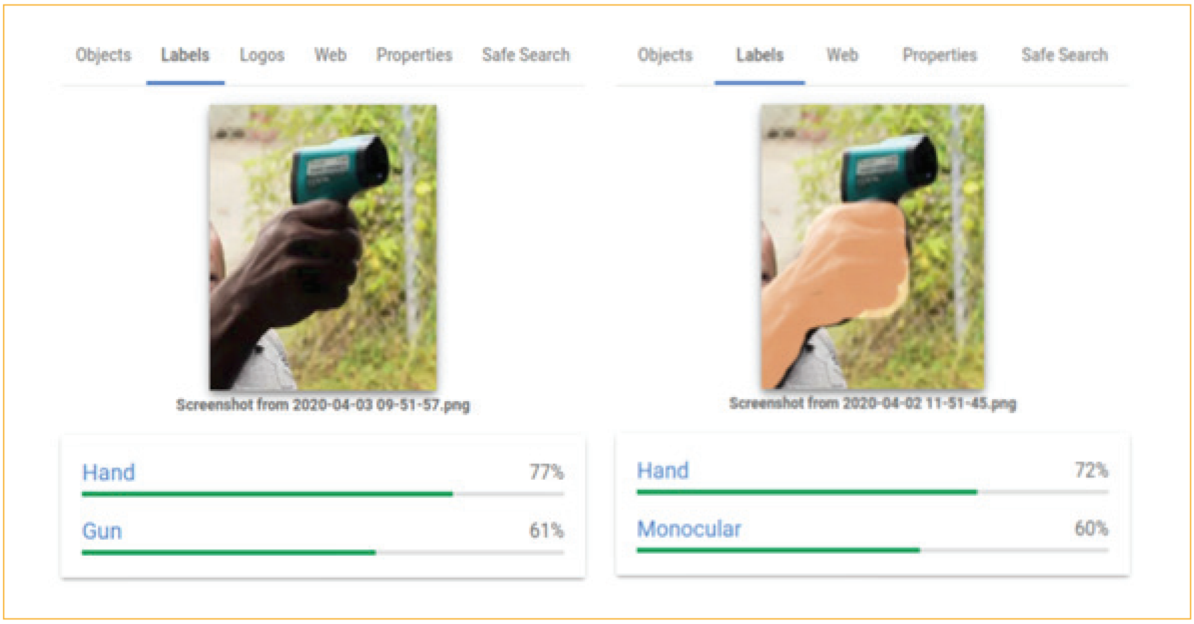


Figure 7 Example of racial bias in Google Vision Cloud

to be of sufficient quality and quantity. First, the training images used should belong to the same distribution as the ones expected in an operating environment. For example, if a system is supposed to recognize objects from CCTV footage, training an AI model with images from professional cameras would probably lead to unsatisfactory results. Moreover, the training data must spread across the distribution of the captured data in an operating setting. Continuing with the example, that means that an image recognition model for an outdoor CCTV camera, should be trained on images made in different weather conditions, time of the day, etc., as long as these are possible operational scenarios.

The amount of good quality data required for model development differs depending on the variety of environment and operational settings and applications of such models (the diversity of possible image and object appearances) and the hardness of the problem. A robust model should cover edge cases too. Additionally, the AI developers must consider possible model biases in various aspects of the model's decisions, whether gender or racial ones for instance, and aim to reduce and altogether remove them. A possible reason for such biases could be not only an imbalance in the number of images, e.g. of people of a certain race, gender or age, but also the source where the data has been collected from. In the past, many issues were reported in which AI models returned inconsistent results for different genders and races. Organisations such as AlgorithmWatch report about and investigate such cases, one example being Google Vision Cloud labelling an image of a dark-skinned individual holding a thermometer as a "gun," while a similar image featuring a light-skinned individual was labelled as an "electronic device".[52]

Lastly, it is essential that the training data is collected in strict adherence to legal requirements and ethical standards. This involves ensuring that no proprietary or copyrighted data is used without proper authorization. Data scraping should only be conducted when it is explicitly permitted by the website's terms of service and relevant legal regulations. Additionally, any personal data should be anonymized to safeguard privacy and comply with data protection laws. These practices are crucial in maintaining the integrity of the data collection process and upholding ethical standards.

Classical image generation techniques

Due to the requirements for good data, computer vision engineers often rely on techniques for augmenting or generating images to increase the quantity and diversity of their datasets. Classically, such methods include geometrical transformations of existing images such as rotation, zooming, shearing and affine; changing of illumination such as brightness, saturation and contrast; and inclusion of artefacts such as occlusions, noise, defocusing and others. Another approach is to place objects of interest on other images, while potentially applying similar transformations to the concrete objects or the whole image.

Original image | **Flip, affine + occlusion** | **Zoom, crop, blur** | **Illumination**

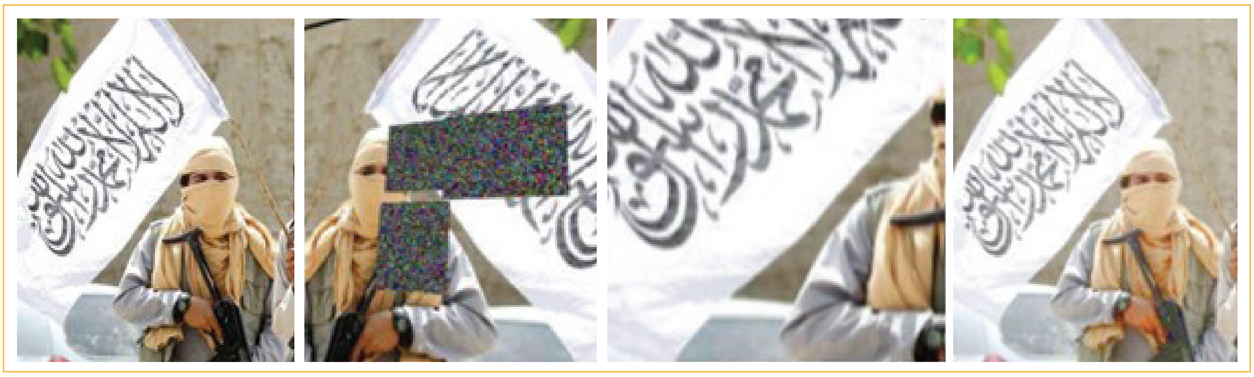


Figure 8 Image Augmentation techniques for dataset enrichment for Content Moderation of radical content (here Taliban flag)

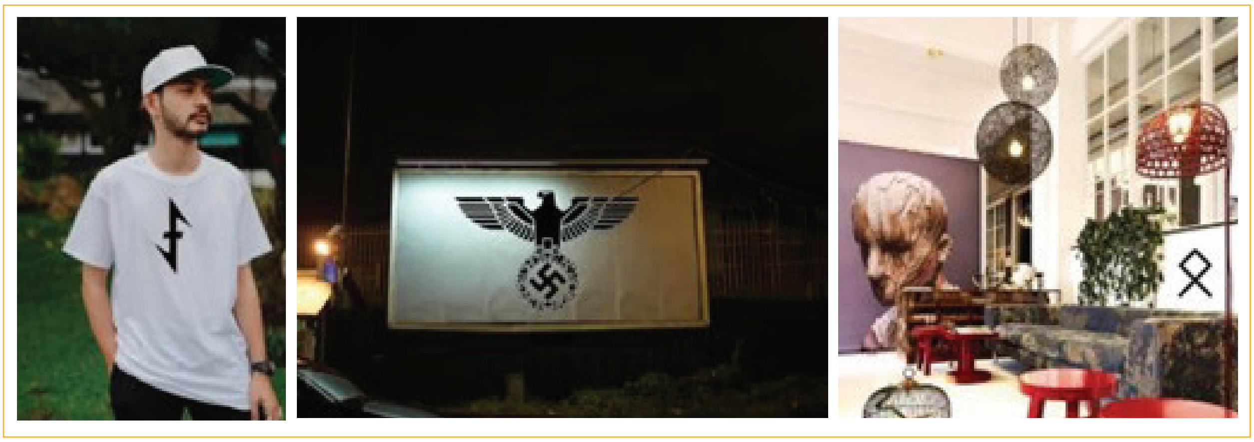


Figure 9 Image Augmentation techniques for dataset enrichment for Content Moderation of radical content (here infamous symbols from left to right: Wolfsangel rune, Reichsadler, Othala (Odal) rune)

GenAI for data enrichment and bias prevention in content moderation

Despite the number of techniques and the diverse results they can produce when applied in combination, the possible appearance of the generated images is still rather limited and often does not resemble real world data sufficiently. The emergence of GenAI, however, introduced new ways of generating synthetic images for model training and evaluation. Model architectural breakthroughs such as Stable Diffusion have made it possible to create realistic looking and diverse synthetic images. These technologies are constantly being improved to create better realism (such as SDXL), speed of image

Figure 10 Result of prompt RAW photo, medium close-up, portrait, solder, black Afro-American, man, military helmet, 8k uhd, dslr, soft lighting, high quality, film grain. altering words related to the race

generation (such as SDXL Turbo), and model training with improved flexibility and reduction in computational requirements (such as LoRA). Additional components for controlling the output have been developed to connect with these technologies such as ControlNet, Inpainting and many others. Furthermore, tools and software for easier and more user-friendly application of the technologies have been created such as A1111 and comfyui for image generation and kohya_ss for training.

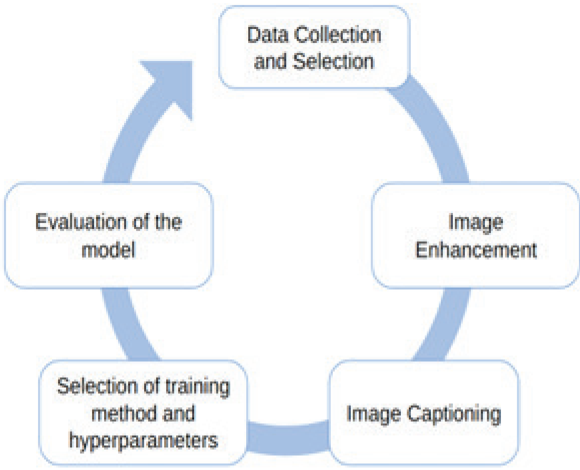
Moreover, there is a great open-source community which, apart from contributing to the fast development, makes in many cases the use of these technologies widely accessible. Developers share their models in platforms like HuggingFace and Civit.ai, allowing others to take advantage of the newest advancements in image generation.

The latest GenAI technologies make it significantly easier to create nearly endless variations of images, thereby expanding training datasets to improve generalisation capabilities and reduce biases. We have been leveraging these technologies within Imagga's tagging system*, tailoring their use to the desired outcomes. In some cases, it is sufficient to allow more flexibility in the underlying models to generate less controlled images. For instance, using text-to-image techniques, we can develop a less racially biased Image Recognition model by generating a diverse set of synthetic images representing people of various races.

In other cases, it is more important to precisely control the output of the image and generate synthetic images where only a specific aspect is altered (e.g., skin colour). By keeping the image label consistent despite these changes, the model is more likely to learn that the desired label is not dependent on those altered characteristics, thus improving its ability to generalise accurately across diverse scenarios.

The next example (Figure 11) illustrates a multistage approach to achieving this, specifically for avoiding racial bias. However, this method is also applicable for expanding training datasets with images representing people of different ages, skin tones, genders, and other characteristics.

First, an AI model is used to segment the pixels of the hand. Next, an inpainting method



targeting the extracted hand mask is applied to generate new content specifically for the hand. During this process, multiple controls are employed to guide the image generation. In this case, Canny is used to define the edges of the hand that must align with the original image, while OpenPose ensures that the model follows the hand structure detected

* For more information see https://imagga.com/solutions/auto-tagging

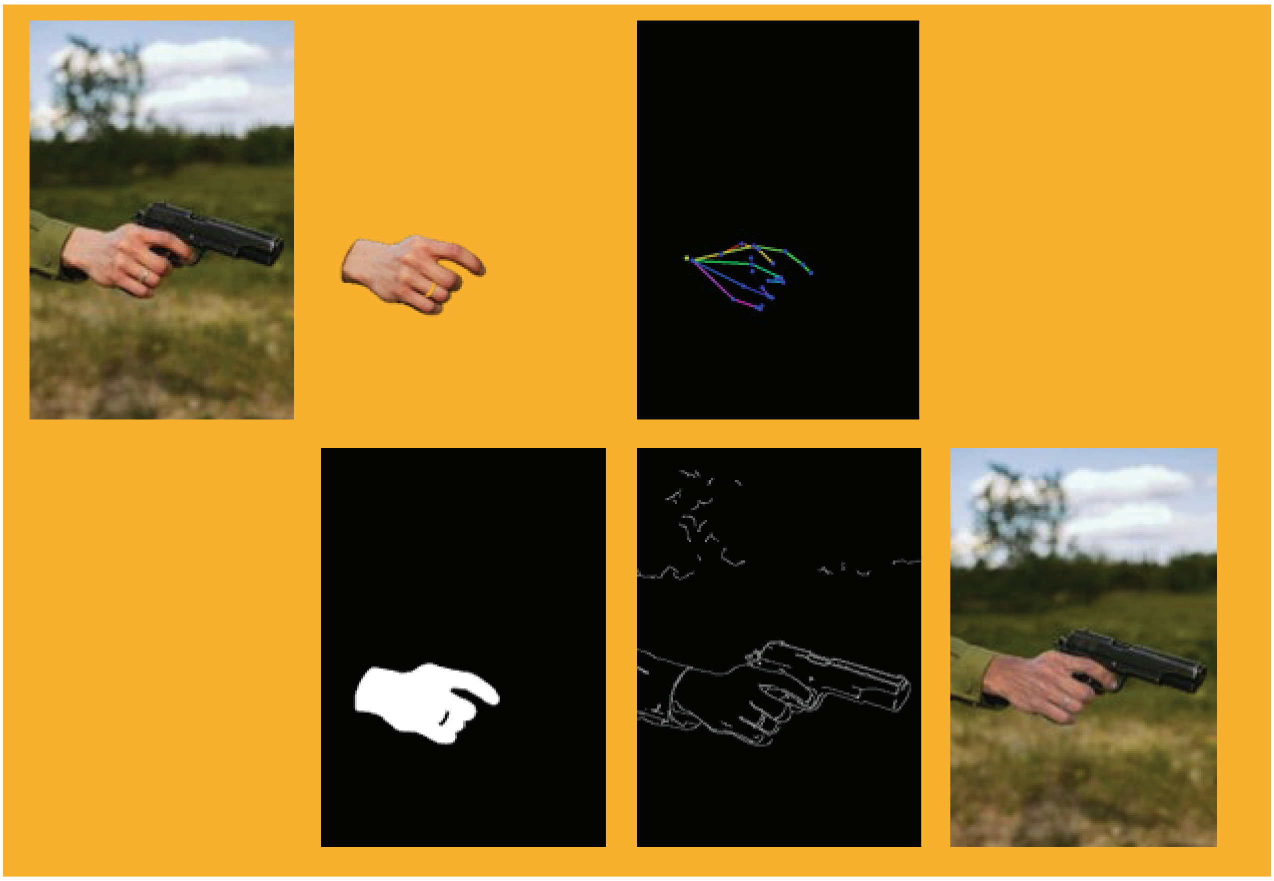


Figure 11 Changing skin colour using multiple steps (left to right: original image; segmentation of hand and mask; result of OpenPose and Canny; resulting image)

in the original image. Additionally, a trained LoRa slider is incorporated into the prompt to control the skin tone, allowing it to be adjusted from white to black.

Another area where advances in GenAI have proven extremely useful is in the development of content moderation systems targeting inappropriate and illegal content. Historically, developing image recognition models for such concepts and objects has been challenging due to the scarcity or unavailability of sufficient training images. At Imagga, we have been leveraging the latest GenAI technologies to expand our training datasets, which has enabled us to create more effective and robust models for identifying radical content. This includes infamous symbols like the Swastika, Totenkopf, SS Bolts, and symbols or flags associated with terrorist organisations such as ISIS, Hezbollah, and the Nordic Resistance Movement. Since this content cannot be generated with open-source models or services, custom training is required to develop these models effectively.

The successful generation of such images involves several steps that require extensive experimentation and hyperparameter tuning, in addition to the training and actual image creation. There are various methods for training a custom GenAI model, each with its own strengths and weaknesses, such as DreamBooth, Textual Inversion, LoRa, and others. Generally, the outlined process needs to be followed to achieve optimal results.

For example, for the development of a Swastika Recognition model, we employed the above approach using approximately twenty images depicting the swastika symbol in different contexts, backgrounds, and formats. The resulting model could generate images containing Swastika with diverse appearances, while maintaining the geometrical characteristics of the symbol.

The new model has been applied both independently and in combination with other models for generating additional image elements like prison cells. Furthermore, it has been successfully used not only in text-to-image but also in image-to-image tasks with Inpainting.

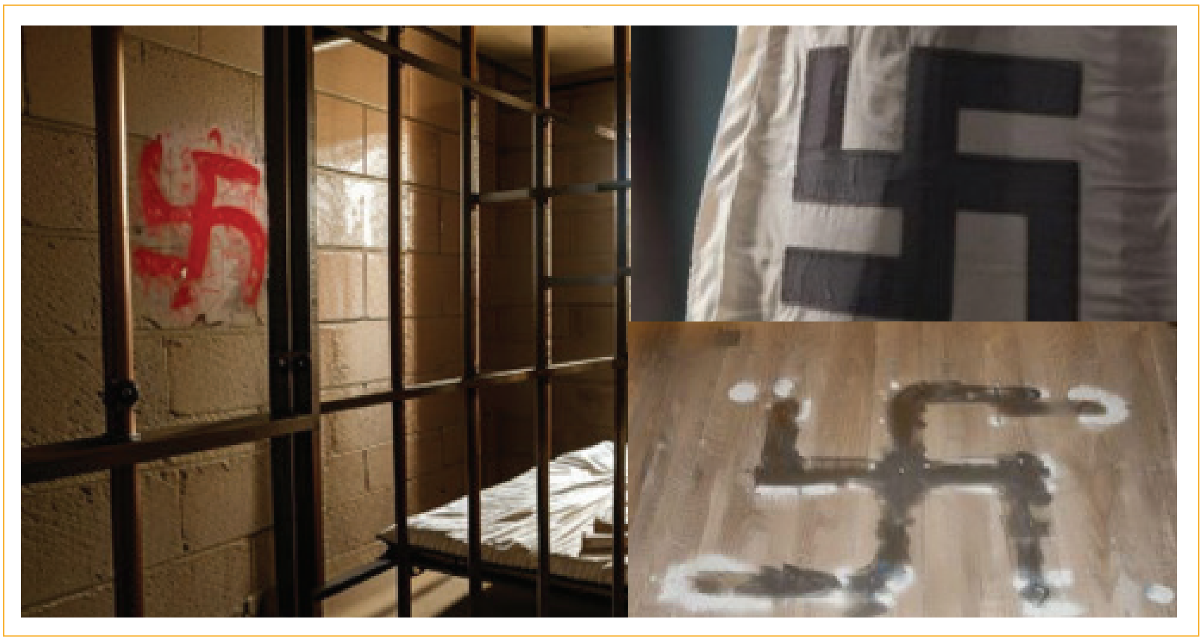


Figure 13 Example of generating additional image elements

The dataset has been additionally enriched with synthetic images produced using other techniques such as Outpainting with LinearArt as a controlling element. This allows them to apply an exact geometrical structure or even text to other backgrounds and alter their style if desired.

Even though training data could be expanded using recent developments in GenAI to improve the performance and generalisation capabilities of Image Recognition models, there are still many challenges associated with these technologies. One major challenge is that the creation of usable GenAI models and the pipeline for image generation requires a significant amount of experimentation. This process is not only complex but also resource-intensive, demanding substantial human effort and computational power.

Furthermore, a critical challenge lies in ensuring that the generated images do not negatively impact the performance of the Image Recognition model when applied to real-world data. Achieving this requires a careful balance between AI-generated images and real-world images. Without this balance, the model may be overfit to the characteristics of AI-generated ones, leading to poor generalisation on actual images. Additionally, AI-generated images or parts of them may need to undergo domain adaptation, whose purpose is to alter the images so that they resemble the real-world images. In some cases, this might also be very challenging, because not all characteristics of an image are visible to the human eye.

Figure 15 Example of synthetic image enrichment using Outpainting with LinearArt to apply geometric structures to varied backgrounds

5

GenAI for Testing and Music Generation in Games

Authors: Rémi Mignot - IRCAM, Christoffer Holmgård - modl.ai

The gaming industry has always been at the forefront of technological innovation, constantly pushing the boundaries of what is possible in digital entertainment. The advent of Generative AI can have a significant impact in transforming game development.

GenAI has the potential to offer game developers powerful tools to automate complex testing processes, ensuring higher quality and more reliable gameplay experiences. Additionally, GenAI can transform the way music is composed and integrated into games, supporting the creation of dynamic and immersive soundscapes that adapt to in-game actions and environments.

This chapter explores the dual role of GenAI in game testing and music generation for games, highlighting its potential to streamline production workflows, enhance creative output, and deliver more engaging and personalised gaming experiences. As with other applications of AI, these advancements come with their own set of challenges, particularly in maintaining the delicate balance between automation and creativity, as well as ensuring the reliability and fairness of AI-driven testing processes.

GenAI for Game Testing and QA

This section deals with current practices and the future potential of applying GenAI methods to game testing and Quality Assurance (QA). GenAI provides a new set of technologies that may prove impactful in the automatic testing of games, though some open questions remain.

First, we provide a brief overview of the state of the art in automated game testing to frame how generative AI may provide novel benefits to existing methods. From there, we describe how GenAI already sees exploratory and production use for automated testing and point to near-term extensions of these use cases. We also account for known and expected concerns and issues related to the methods.

From there, we point to potential future uses of GenAI for game testing and QA and conclude by summarising the differences between current methods, near-term GenAI methods, and long-term GenAI methods.

State of the art in AI-driven automated game testing

Automated testing of video games using artificial intelligence methods is still an emerging field in the game industry, where the vast majority of game QA is done manually and most often by leveraging outsourced resources, such as remote teams, generalist services vendors, or specialised QA vendors.

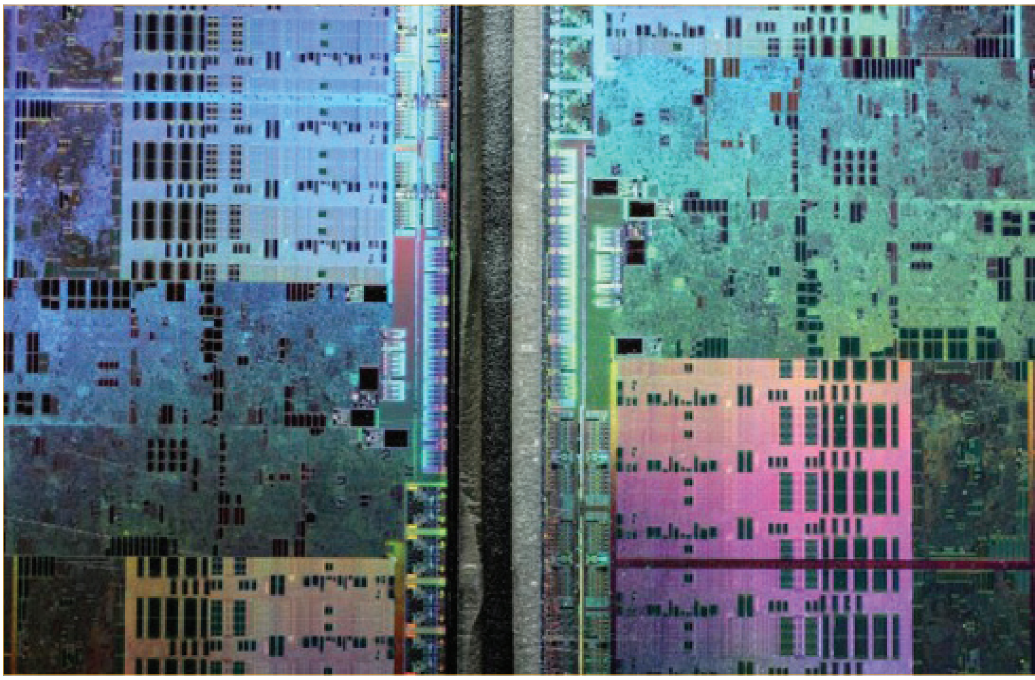
In recent years, a movement toward automated testing in game development has grown, partly informed by the observed success of such practices in other kinds of software development. The topic has received increasing interest at premier professional game developer venues, such as the annual Game Developers Conference[53], GameQuality.org[54], and Game Quality Forum.[55]

Automatic testing of games does not necessarily imply using AI for testing processes. Many practices and processes in automatic testing, also for games, can be achieved through unit tests and continuous integration testing, sometimes coupled with deterministic programming for executing in-game behaviour in lieu of a human player. The approaches, however, reach a natural limit when applied to game testing due to certain inherent aspects of games as software products: Most games execute as real-time simulations, with no reasonable ways of pausing or rewinding the simulation timeline.[56]

They typically also feature elements of randomness and emergent interactions of their components, either by design or through imprecision from rounding errors in internal systems such as physics engines, usually necessary due to the complexity of the simulations and the performance requirements of games.

Additionally, most games are highly multimodal experiences, encompassing logical and dexterity challenges that combine narrative, auditive, visual, and haptic content. Testing whether these interacting components and content all work as intended and produce the intended experience is a task that has typically defied full automation. Still, some headway had been made through AI and machine learning techniques to operate parts of the process, namely operating the game itself in the way that a player would by playing it. The approaches have improved efficiency for some parts of the "testing pyramid" in games, automating repetitive tasks for QA testers. Still, these machine learning-based approaches only address part of the testing pyramid, as they may struggle with or cannot reason over complex or novel content in the games being tested.[57]

In the next section, we outline how GenAI addresses or points to a path to address these topics.



Fritzchens Fritz / Better Images of AI / GPU shot etched 5 / CC-BY 4.0

Generative AI in game testing

Game testing and QA fundamentally comprise two parts interlinked in a single process: 1) Playing the game to elicit errors and analysing the game as it is running, and 2) analysing generated data after a run to identify issues from a generated data set. These two parts may occur at the same time or separately.

So far, most headway has been made in the use of GenAI for using Large Language Models (LLMs) to Vision Language Models (LLMs with attached visual adapters that allow them to consume visual information in the form of pixel renders). Several companies in the space of automatic game testing have started offering early solutions that allow game developers to submit textual or visual information for analysis.

LLMs may consume textual information to identify issues in large volumes of logging data from executed game runs, identifying issues and proposing likely causes and remedies to the QA specialists and programmers interacting with the models. This approach saves human staff the time and effort of meticulously inspecting the generated data sets and deciding the next steps.

A second in-market approach is to use Vision Language Models (VLMs) to inspect still frames or recorded videos from game playing to detect visually observable errors. This may be useful for functional testing of video games, similar to non-GenAI-enabled AI-driven testing. A strength of using GenAI Vision-Language models, however, is that the models enable the identification and description of issues encountered for later consumption by human QA specialists or other parts of the development team. This approach has the potential to radically change recurring, manual, labour-intensive game testing tasks such as checking for rendering issues/glitches or checking for issues in game Graphical User Interfaces (GUIs).

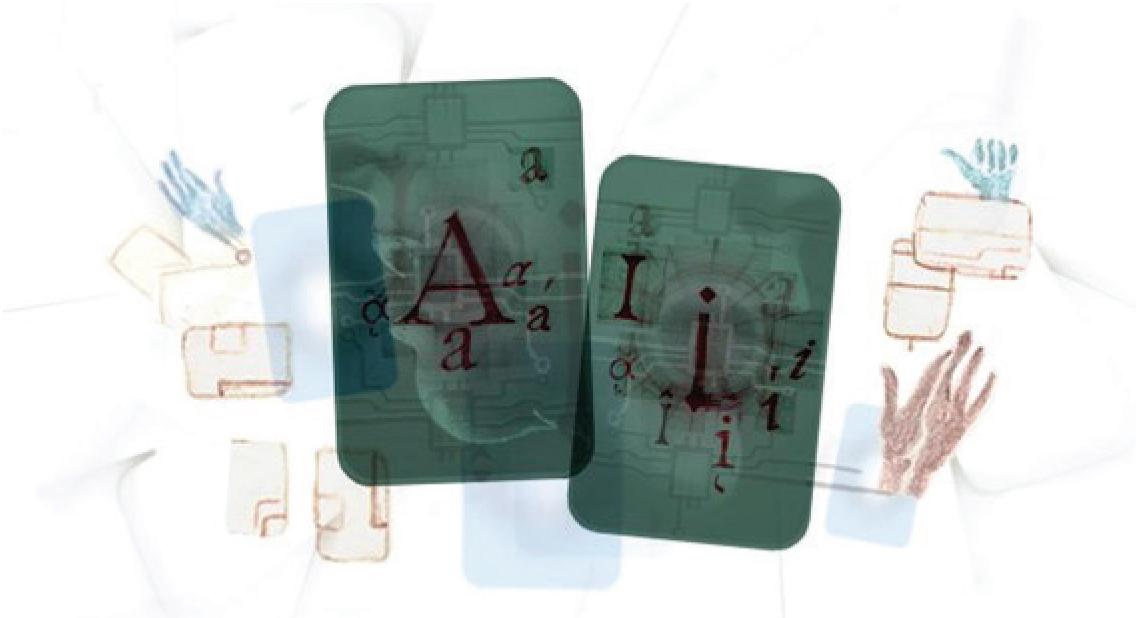
Future applications of Generative AI in game testing

While GenAI in game testing and QA has been chiefly used in the analysis side of the process so far, research and development work is ongoing on leveraging LLMs and VLMs at run-time for game playing. Several commercial actors have released videos with proof-of-concept game-playing agents that operate games from textual instructions provided by human operators, parsed by LLMs[58], or by reading screen information and making gameplay decisions using VLMs to analyse the content.[59] While these approaches are still in the early stages, they may, in time, provide a more general way of enabling AI-driven automatic testing for games than the methods described above.

In summary, the use of GenAI for automating game testing and QA is still in its infancy in the games industry. Still, multiple actors are innovating in the space, bringing concrete solutions to the market. Regardless, any adoption of GenAI methods at scale in game testing and QA will depend on methods working 1) at low cost, 2) with acceptable precision, and 3) delivering models and processes that can be accepted and integrated into existing production processes. Overall, implementing GenAI approaches into games industry processes may face challenges from the process change resulting from automation and a broader discussion about the ethical training and use of large foundation models.

GenAI for Music Generation in Games

In this section, we discuss the potential of utilising GenAI and LLMs for music production in the video game industry. We first focus on GenAI tools that allow musicians to create music in a new way, then on GenAI with LLMs for an automatic music generation based on textual queries. Thanks to our own insights, and using outputs of surveys and interviews with professionals of the sector, we try to answer the following questions: What are the current practices in the creation of background music for video games? What do GenAI and LLMs bring to the table for music creation? What are the limitations of today's automatic music generation for video games? How can GenAI and LLMs be used for video game music today? What are the next challenges for integration into video game workflows?



Alina Constantin / Better Images of AI / Handmade A.I / CC-BY 4.0

Current practices in the creation of music for video games

Even if the stakeholders of the video game industry are used to new technologies, the production practices of music for video games seem more traditional. After having defined the musical ambiances of the game (mood, genre, instruments, etc.), the video game managers employ human musicians to compose and to record the music.

For video games, music creation has additional constraints: first, the background music should not attract the player's attention, its role is to enhance the atmosphere of the game, leaving the focus on the action and the story. Moreover, in some cases, the music and the sound design are made in collaboration, in order to make a unique soundscape which merges music and sound events (sound designers are usually also musicians). Another important point in video games is the interaction with the action of the game. For example, if the story goes from a calm moment to a frightening one, the music needs to change instantaneously with a coherent change. This point is probably the most difficult and technical aspect of music integration in video games.

To deal with interaction, two methods are commonly used. The most straightforward one is the horizontal method, which consists of playing the music linearly, and in changing the played piece when needed, for example when the action changes. To create a smooth change, music transitions are prepared by the composer. Depending on the game, this first method can be well suited, otherwise the other method is used. With the vertical method, the composer provides several recordings for each composition. The recorded music is split into different layers (corresponding e.g. to different groups of instruments) and it is given with different variants and moods (calm, energetic, etc.). Then, the music integrator of the development team programs the synchronisation of the game and the music.

For example, the composer and sound designer Yann van der Cruyssen says for the music of the game Stray: "Rather than a fixed linear planning, the music is a mix of several playlists that use conditions, timers, and priorities to adapt to the player's style as much as possible. Each tune is either strongly associated with a given action or randomly fades in after a certain period of silence or something in between".[60]

Additionally, in the past few years, procedural tools are being more and more used to enhance the interaction between music and action. Such an example is the audio engine Metasound of Unreal Engine[61]; rather than using machine learning, such tools are based on programmed rules and constraints and make the work of the music integrators easier. Such tools still require the intervention of human composers.

Advantages and limitations of GenAI and LLMs for music creation

As it has been reported for several years during the Sónar festival discussions[62], GenAI is more and more used in music, especially for digital and experimental music. GenAI models offer new ways to shape sounds and create music. The number of available GenAI models continuously grows. For example, a well-known tool in the music community is RAVE.[63,64] It allows for a real-time timbre transfer, i.e. it can transform an instrument to another one (percussions to voice, for example) or it permits a manipulation of the sound through the latent space (a high-level audio representation). Another example of a GenAI tool for music is NOTONO.[65,66] Based on a GenAI model, it proposes to musicians a Graphical User Interface (GUI) to finely sculpt sounds. For music accompaniment, a series of GenAI tools[67] has been proposed to generate bass lines either through a GUI and controllers[68] or based on the main music tracks given as input.[69]

The common features of the mentioned tools are:

- Controllability: Expert musicians value control over their audio production. By design or through GUIs, most GenAI tools allow musicians to oversee and manage sound and music generation.
- Easy integration into standard digital music workflows: In digital music creation, musicians typically use a chain of devices, software, and plugins connected together. Integrating GenAI tools into these workflows poses no issue for musicians who are accustomed to adopting new technologies.

Generative AI with LLMs goes even further by automatically producing complex music mixes based on textual queries solely. In a prompt, the user gives to the model many aspects of the wanted music, such as the genre, the instruments, the mood, the harmony, and possibly asks for music similar to a given artist. Among the first developed tools, we can cite MusicLM[70,71], and MusicGen.[72] Even if their music and sound qualities are very poor, these models made waves. Then, some companies such as Suno[73], Udio[74], and Beatoven[75], developed new models to achieve outstanding results. Not only the sound quality has been significantly improved, but also the expressivity of the instrument and voice tracks are more convincing. Even the harmony, rhythms, and arrangements are sometimes richer and more complex than many current popular music.

Nevertheless, even if the sound quality has significantly improved since the first models, the generated sounds and music still have some shortcomings. For example, with all current models, the transient parts, such as percussion and note attacks, are not yet efficiently reproduced. Additionally, the compositions and arrangements can sometimes be questionable: the alternation of moments of tension, rest, and climax is not always coherent. These models still require improvements, both at a micro level (e.g., transients) and at a macro level (overall composition and arrangements).

The current products for automatic generation of music from textual queries are mainly oriented to people without skill in music. Not only does the user have no fine control on the produced music, but also such models usually do not understand specific music terms, such as the time or the key signatures. Contrarily to what we could expect, composers for video games like audacity and originality in music. For example, Kenta Nagata used the unusual 11/8 rhythm signature for a music piece of Mario Kart 64.[76] This produces an abnormal rhythm which may unsettle listeners, but the author cleverly arranged the composition to avoid destabilisation. Currently, with automatic music generation models there is a risk of obtaining music with too little originality.

Finally, and more specifically for video games, GenAI with textual queries could offer significant benefits in automatically producing background music, potentially replacing the need for human musicians. However, in their current state, these advanced tools are not yet directly usable for video games due to two major constraints: controllability and interaction.

- **Controllability**: According to the current state of the art, musicians and sound designers don't have enough control over the produced sounds and music. For example, when the generated music is not satisfactory, the user can only modify the text or relaunch the query to get a new random proposition.
- **Interaction**: With such tools, no solution is proposed to manage the interaction between the generated music and the game. Moreover, AI models cannot anticipate music transitions as human composers do.

Current uses of GenAI and LLMs for video game music

The use of GenAI models by musicians is obviously not a challenge, at least for those who are used to advanced and experimental technologies. Many tools, such as RAVE or NOTONO, can be easily integrated into the usual workflow as new plugins or software in the processing chain. But models of automatic music generation from textual prompts are obviously a concern because they are not intended for skilled musicians.

On the side of video game producers, because of the limitations mentioned above, the use of today's models of automatic music generation can be possible for two limited purposes:

- They can quickly provide music

placeholders to make audio mood-boards, to be used in the first prototypes, and possibly to give examples from which the human composer must draw inspiration.

- In the case of small video games, with fast development and with a restricted budget, these models can quickly and effortlessly generate music. The trade off is that the quality is lower, and the cohesion with the game (interaction, atmosphere, etc.) is weaker.

Some musicians already use GenAI to experiment with new ways of music making, but the existence of automatic music generators is a concern for them because they lose control. For video game producers, the use of such tools is possible, but some obstacles must be removed. These obstacles are mainly related to the level of controllability and interaction with the model and the issue of limited resources.

To conclude, these new technologies also raise a more general issue: since these models imitate the existing music, they are unable to generate original music and to give rise to new musical trends. Even if video games represent a tiny part of musical creation, they still participate in it. Note that for some video games, the musical creation is worthy of a Hollywood production, see for example Final Fantasy VII[77], God of War[78], Civilization VI[79] and Assassin's Creed Mirage.[80] As with the cinema, video games allow players to hear musical styles that they would not listen to on their own, and thus help to expand their musical culture. For example, everyone who has watched "2001: A Space Odyssey" has heard the music of Strauss and Ligeti at least once in their life.[81]

6

GenAI for Music Co-Creation

Author: Artur Garcia - Barcelona Supercomputing Center (BSC)

The development of generative tools has emerged as a major application of Machine Learning, and generative applications to music are no exception.

From initial generative models that appeared just some years ago to current Neural network (NN) architectures, the scale at which revolutions appear in the field is measured by months or even weeks. This scenario is having a great impact on artists who are potentially the final users of this technology, but who observe these developments with a concern about the impact these advances will have on their own careers.

At the moment of writing this chapter (July 2024), every major actor in the AI spectrum, together with a collection of new companies, are pushing the use of AI for music creation to limits beyond any expectation set a few years ago. This collection of new applications addresses a number of tasks not restricted to music creation, but also including music classification, recommendation engines, and music conversion. Generative models have become ubiquitous in any musical creation environment, with users having access now to tools that were restricted to professionals a few years back. This scenario translates into professional capabilities and output from really modest conditions, democratising the creative process.

AI is currently transforming many aspects of the musical industry and market and is having an impact on every aspect of the music industry. Copyright laws, artists ownership, music styles, and streaming platforms, all will have to adapt to a rapidly changing landscape where it is difficult to predict even which new topics will become central in the discussion in the next months. Every new tool has the potential to revolutionise the field.

Yutong Liu & Kingston School of Art / Better Images of AI / Talking to AI / CC-BY 4.0

We address here the recent advances in the field, presenting a chronology of advances in the preparation of generative models. We discuss the changes happening in the industry, and potential implications that these developments will have in the community of creators. By focusing on creators as the main users of this technology, we highlight the double edge sword character of this technology.

Recent evolution of the field (last 5-10 years)

Recent generative models are providing an opportunity for creators to access professional grade tools to enhance their production capabilities and reach a wider audience. Technology, in music production, has meant a democratisation of content production. From a technical point of view, music is a particular case of data analysis related to series prediction, which is a central topic in data science and machine learning. Recent methods proposed for series prediction like recursive Neural Networks have been extensively tested on music generation, including Recursive Neural Networks (RNNs) and Large Language Models (LLMs).

Focusing strictly on the applications to music generation, the pre-trained neural networks have provided impressive results and have been applied to multi track music composition. However, the long-range correlations of music, meaning that relations between musical notes are established at different scales, make this problem computationally hard. It is due to this limitation that the evolution of the field has followed these steps:

- Initial models for music processing based on MIDI files using RNNs.[82]
- Developments of text-to-speech models generating audio tracks using Convolutional Neural Networks (CNNs).[83]
- Generative adversarial models trained on audio files.[84]
- Text prompt for music creation, with models for music generation as large as 3.3B parameters.[85]

Following the developments of the field introduced above, several ongoing projects are producing a large collection of tools for end users. While some of these tools are presented as final products, many of them are presented in the form of pre-trained models that users have to manage to install and use. This may pose a restriction for non-expert users with little technical training.

In addition to low level developments producing new generative models, a range of top-level tools allow the use of these models in non-technical environments. In AI4Media's Use Case 6 for Music co-creation, Barcelona Supercomputing Center (BSC) has led the implementation of an integration tool that allows users to use in an effortless way a large collection of different generative models from different sources. This effort unifies the generative capabilities removing the task of integrating the model to a specific platform, an effort out of reach for many users. Much in line with these efforts, large industrial players have developed their own tools, in some cases in collaboration with renowned artists such as Google MusicLM or Deepmind's Lyra. Having developed this collection of tools, these companies can target from this moment the large market of end users interested in these advances, adding this technology to the portfolio offered in their current services.

Industrial impact: Current models and applications

The intersection of AI research and music has produced several use cases where Machine Learning operations allow the processing of music in different formats. Among these different tasks one finds:

- Music enhancement, where an audio track is augmented with additional music.
- Music composition, where a model creates the score of a composition.
- MIDI generation, where MIDI files are converted into audio files.

The field of audio and music creation has integrated recent advancements in ML techniques, but as a result, it is now facing the same set of challenges that arise from using these tools. Among these, the quality of musical content raises questions about authorship and copyright. On one hand, models can reproduce with great fidelity compositions by given authors. On the other hand, the utilisation of copyrighted material in the training process leaves the authors unprotected to claim their ownership. These legal implications are becoming central to AI development.

The music industry is adapting to a rapidly changing landscape due to the rise of AI technology. Current companies well established with products widely used, have to adapt rapidly to absorb these advances and provide them to their users. Digital Audio Workplaces like Ableton, Logic Pro, or even Adobe already offer integration of novel tools -sometimes developed by third parties- to their existing line of products. This approach allows musicians used to a particular workplace to integrate the new tools easily in their workflow.

However, the emergence of AI has also resulted in a rise of new companies with dedicated products. Companies like Soundraw, udio.io, and sunio.ai offer powerful music composition tools on standalone products. Despite their recent apparition, these new models are already well established in the music community and represent the forefront of a changing landscape of novel applications and models.

Finally, major industrial tech companies are actively developing general AI tools, with applications to music production. This can be perceived as a reaction to the market movements and the potential risk of losing ground with novel companies. These companies -well established at the cutting edge of ML and AI research- are actively releasing projects showcasing state of the art capabilities of these techniques. The following projects, all by major AI labs in large tech companies, are a good example of these activities:

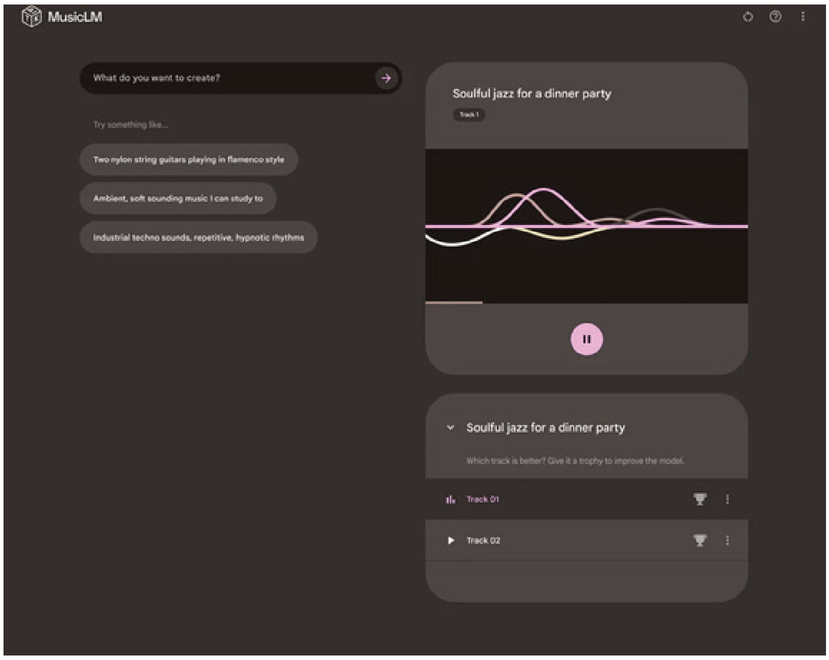


Figure 16 MusicLM is a model generating high-fidelity music from text descriptions. It is Google's effort to use LMs for music generation.

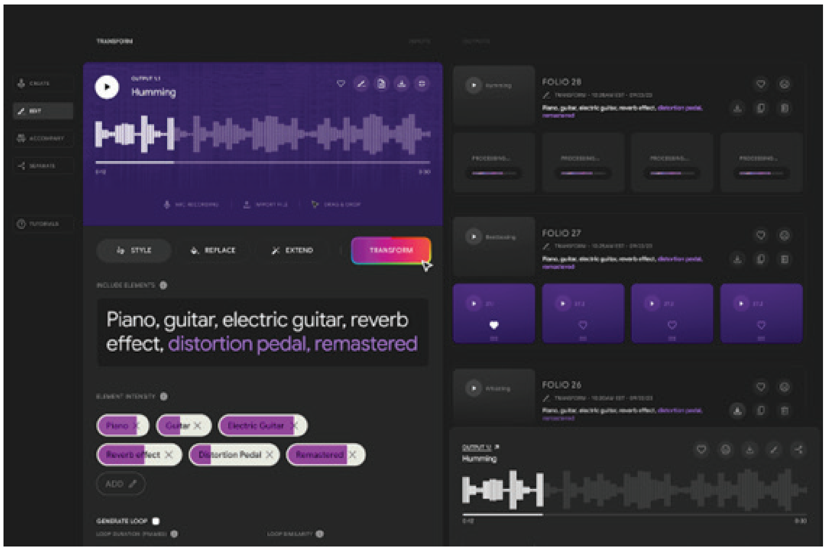


Figure 17: The Deepmind Lyria model has the capability to generate high-quality music with instrumentals and vocals, performing transformation and continuation tasks, and giving users more nuanced control over the output's style and performance through its novel user interface.

- Deepmind: Together with the Lyria project[86] presented above, Deepmind has also introduced models for the generation of audio for video (V2A) coupled to video models such as Veo.[87]

- Meta: Meta introduced Audiocraft[88], which includes AudioGen and MusicGen, two generative models developed as a single autoregressive Language Model (LM).
- Microsoft: Current work in Muzik[89] includes models for music understanding and generation.
- OpenAI develops and openly distributes Jukebox[90], a generative model for music including rudimentary singing -as raw audio- in a variety of genres and artist styles.

Near term future: Current and future use in real cases

The advances in the field of music composition and generation have identified several tasks for which a trained model can provide professional grade results. However, current limitations in the model capabilities and the technological requirements may prevent the complete use in any scenario that professional musicians may be interested in.

An example of current limitations of AI models is their usage in real live performance. Most of these tools are intended for music generation and production in the studio, while their uses as effects in real life performance are limited. This is the result of the requirement to run large models able to process high quality audio with professional grade latency.

In collaboration with local artists in the context of the AI4Media project, BSC has led the development of music processing tools for live performance. In this prototype project, the generative model is based on physical properties of the human body to generate sounds. This approach allows a massive reduction in the number of parameters required, while producing realistic transformations in real time. This is a step in what may become a major application in the short term for artists, combining small models with the powerful adaptability of the large model.

This is a potential usage that may advance dramatically in the upcoming years, with smaller models able to run in real time but with great processing capabilities (as an example, META's MusicGen model includes a small version with only 300M parameters). Additionally, these models may focus on particular aspects of music production like voice processing, orchestration, or improvisation.

In summary, the number of new models, generative tools, companies, and initiatives are an indicator that the maturity of the intersection between music and AI is rapidly advancing. We have reviewed some of the advances in music composition that are currently released by new companies, both as generative models or full development platforms, and the efforts with which major companies are leading the field. However, a true moment where musical tools are ubiquitous and extensively used by non-musicians may not have arrived yet, and one may consider that a revolutionary moment is still to appear. In this way, the perception among musicians is that upcoming changes are near, due to the rapid advances of the field combined with the perception that in other areas -such as image generation- the results of current models are on a higher level. Finally, while for many artists these advances are perceived as a big opportunity, other users express concerns about the impact these will have on their careers.

7

Challenges of Integrating GenAI in the Media Industry

Integrating GenAI into the media industry offers significant potential but it also presents a series of challenges that span ethical considerations, technical difficulties, operational constraints, and industry-specific concerns. This chapter briefly discusses the main challenges that are associated with the use and integration of GenAI technologies into the different media processes that were presented in the above sections.

One of the most pressing issues is the **ethical and bias concerns** inherent in AI technologies. GenAI models rely heavily on the data they are trained on, and if this data is biased or skewed, the AI will perpetuate these biases in its outputs. This is particularly troubling in areas like content verification (discussed in Section 2), video production and content automation (discussed in Section 3), and content moderation (detailed in Section 4), where biased outputs can lead to misinformation or unfair race, age, and gender representations. Furthermore, the phenomenon of **AI hallucinations**, where the model generates plausible but incorrect or entirely fabricated information, poses a significant challenge. This issue is especially problematic in news production and fact-checking and content verification, where rigorous human oversight becomes necessary, complicating the integration of AI into media workflows. The opaque nature of AI decision-making processes also makes it difficult to understand how conclusions are reached, which can erode trust in AI systems and complicate efforts to hold AI accountable for its outputs. As such, the nature of some AI algorithms raises the problem of the **"black box society"**, a term coined by Brooklyn Law School professor Frank Pasquale to express the notion of a networked society based on opaque, non-transparent algorithmic systems.[91] These concerns raise the issue of **transparency and explainability of AI systems**; how can we make AI more transparent and explainable, and is this fully possible? The issue of algorithmic transparency or opacity is also closely connected with other issues, such as responsibility (an ethical concern over who should be responsible over automated decisions), liability (a legal concern about who should be held liable for unlawful consequences), and explainability (as a philosophical concern over the very nature of explanation and the human decision-making process).

Technical challenges further complicate the integration of GenAI in media. Maintaining **consistency and quality** in AI-generated content, particularly in video production and content automation (as highlighted in Section 3) as well as in music production (discussed in Sections 5 and 6), is a complex and resource-intensive task. For instance, ensuring that AI-generated videos maintain coherence across frames, or that music compositions align with human expectations, requires sophisticated algorithms and significant computational power. This **demand for computational resources** (a common issue identified throughout this white paper) not only raises environmental concerns due to the energy consumption of data centres but also creates barriers for smaller organisations that may lack the infrastructure to deploy these technologies effectively. There is also a **demand for increased controllability,** especially in the domains of music generation (as detailed in Sections 5 and 6) and content moderation (as highlighted in Section 4), to enable users to better control and refine the outputs of GenAI tools.

On an **operational level**, integrating GenAI into established media workflows can be highly disruptive. Media organisations must carefully balance the benefits of automation with the need to maintain **human oversight**, particularly in creative processes like video production (Section 3) and music composition (Section 6), as well as in editorial tasks related to content verification and news media production (Section 2). The risk of **vendor lock-in** is another concern, as relying on proprietary AI solutions can limit an organisation's flexibility. Alternatively, using open-source models requires a substantial investment in expertise and infrastructure, which can be prohibitive for many companies. The adoption of GenAI also necessitates **significant changes in organisational processes**, particularly in quality assurance and game testing (as discussed in Section 5). These changes can meet resistance from staff who are accustomed to traditional methods and may not fully trust automated solutions.

Legal compliance is another critical area of concern. The use of AI in media production involves handling large amounts of data, raising serious issues about **data privacy and security.** Ensuring compliance with data protection regulations, especially when using third-party AI services, is crucial but challenging. Additionally, **intellectual property and copyright issues** come to the forefront as AI models trained on copyrighted material may inadvertently reproduce or mimic existing works, leading to potential legal disputes. This is especially relevant in music generation (discussed in Section 6), where the fine line between inspiration and replication must be carefully managed to avoid infringement.

Industry-specific challenges also play a significant role in complicating the integration of GenAI. In content verification (discussed in Section 2), the rapid spread of misinformation, exacerbated by the rise of synthetic media, presents a major hurdle. Traditional verification methods are increasingly inadequate, and while GenAI offers new tools to combat this, they are not foolproof and require ongoing development to stay ahead of malicious actors. Content moderation faces similar challenges, particularly in identifying and managing harmful content like extremist symbols or disinformation (Section 4). The effectiveness of these AI models depends heavily on having large, high-quality datasets that accurately represent the conditions in which they will operate. Moreover, ethical concerns about potential biases in moderation decisions remain prominent.

In creative industries like music and game production (discussed in Sections 5 and 6), the challenge lies in **using GenAI to enhance, rather than replace, human creativity.** AI tools must be integrated in a way that supports artists and developers without compromising the originality and quality of the final product. Balancing the powerful capabilities of AI with the human elements critical to media production is essential to overcoming these challenges.

The integration of Generative AI into the media industry involves navigating a complex landscape of obstacles, but these challenges also present opportunities for innovation. Addressing them requires a thoughtful approach that prioritises responsible implementation, ensuring that the benefits of GenAI but also generally AI technologies are realised without sacrificing quality, fairness, and transparency. Understanding and addressing these challenges—whether in content verification, video production, content moderation, or creative processes like music and game development—is crucial for developing responsible, compliant, and efficient Generative AI applications that can be widely adopted by the media industry.

8

Conclusions

The integration of Generative AI technologies into the media industry marks a transformative moment, offering new opportunities to enhance creativity, streamline workflows, and improve the accuracy and efficiency of content production, verification, and moderation Throughout this white paper, we have explored the multifaceted applications of GenAI, including its use in content verification, video production, content moderation, game testing, and music generation. While the potential benefits are substantial, so too are the challenges that accompany the adoption of these technologies.

Key takeaways

Key takeaways from the above discussion and analysis include:

GenAI's transformative potential: GenAI has the power to fundamentally transform media production processes by automating tasks that were once labour-intensive, such as video editing, music composition, and content moderation. Its ability to generate high-quality content quickly and at scale can significantly enhance the productivity of media organisations.

Ethical and bias challenges: The deployment of GenAI in media can be hindered by ethical concerns, particularly related to bias, transparency, and explainability. These challenges underscore the need for responsible AI development, where transparency, fairness, and accountability are prioritised.

Technical hurdles: Maintaining consistency and quality in AI-generated content remains a complex challenge, particularly in creative domains like video production and music composition. The demand for computational resources, coupled with the need for more controllable and explainable models, highlights the ongoing technical development required to fully realise GenAI's potential. The problem of AI hallucinations is also an important hurdle that negatively impacts the wider adoption of Gen AI tools.

Operational disruption: Integrating GenAI into existing media workflows requires significant changes in processes and organisational culture. The risk of vendor lock-in, coupled with the need for extensive retraining of staff, can create barriers to adoption. However, with careful planning and investment, these challenges can be mitigated.

Legal and compliance issues: The use of GenAI raises important questions about data privacy, intellectual property, and copyright. Media organisations must navigate these legal landscapes carefully to avoid potential liabilities and ensure that AI-generated content adheres to all relevant regulations and standards.

Industry-specific considerations: Different sectors within the media industry face unique challenges when adopting GenAI. For instance, content verification must contend with the rapid spread of misinformation, while music and game production must balance the use of AI with the need to preserve human creativity and originality.

Recommendations

As the media industry continues to evolve, the role of GenAI will undoubtedly expand. In the near future, we can expect AI-driven tools to become more integrated into daily media operations, assisting not just with content creation, but also with enhancing audience engagement and improving the overall user experience.

However, the successful adoption of GenAI will depend on addressing the ethical, technical, and operational needs and challenges identified in this white paper.

By translating these challenges, needs, and pain points to recommendations, this white paper highlights several areas where further research and innovation are crucial. The development of more **transparent and explainable AI models** is of paramount importance, especially as these technologies are increasingly integrated into sensitive applications like content verification and moderation. There is also a need for **advances in bias mitigation techniques,** ensuring that AI models are trained on diverse, representative datasets and are rigorously tested for fairness across different demographic groups. Additionally, researchers should focus on i**mproving the controllability and consistency of AI-generated content**, particularly in creative fields like video production and music composition, where quality and coherence are critical. **Developing lightweight, resource-efficient models** that can operate effectively in real-time applications without sacrificing performance is another key area of exploration. Finally, **collaboration with industry stakeholders** is essential to ensure that AI technologies are not only technically robust but also align with the practical needs and ethical standards of the media industry.

Going forward, it is important for all stakeholders involved in the process to consider and address the following key points and issues:

- **Investing in ethical AI development**: Media and research organisations should prioritise the development/integration of AI technologies that are transparent, fair, and accountable. This includes implementing rigorous testing for bias, ensuring transparency in AI decision-making processes, and adopting explainable AI techniques.
- **Adopting a collaborative approach**: Collaboration between media companies, AI developers, researchers, and policymakers is essential. Joint initiatives can help to advance AI technologies in ways that align with the ethical standards and practical needs of the media industry.
- **Enhancing AI literacy**: Invest in training programs to upskill media professionals, enabling them to work effectively with AI tools. Building AI literacy across the organisation will help to overcome resistance to change and ensure that GenAI is used to its full potential.
- **Establishing robust data governance**: Ensure that all data used for training and operating AI models is ethically sourced, secure, and compliant with data protection regulations. This includes implementing strict data governance frameworks that cover data collection, storage, and usage.

- **Promoting industry-wide standards**: Work towards establishing industry-wide standards for AI use in media, particularly in areas like content provenance, watermarking, and AI-generated content labelling. These standards will help to build trust with audiences and users and ensure the responsible use of AI technologies.

As the media industry embraces Generative AI, there is a unique opportunity to thoughtfully integrate this technology into various aspects of content creation and production. By approaching this integration with a focus on ethical standards, transparency, and collaboration, GenAI can become a valuable tool that enhances both the efficiency and creativity of media professionals. The challenges are significant, but they can be managed with careful planning and a commitment to responsible development. We hope that the insights and recommendations in this white paper provide practical guidance for media stakeholders and AI researchers, helping them to develop GenAI technologies that can serve the needs of the European media industry, maintaining at the same time the integrity, standards, and values that are fundamental to the media industry and the European society.

Glossary

Below we provide a glossary of technical and non-technical terms frequently mentioned throughout the white paper in alphabetical order.

Term	Definition
AI Hallucinations	Instances where an AI model generates plausible but incorrect or entirely fabricated information.
Algorithmic Transparency	The degree to which the operations of an AI system can be understood by humans, addressing the "black box" problem. See also "Black-Box problem" below.
Augmentation	Techniques used in AI to increase the diversity of training data by applying transformations like rotation, zooming, and brightness adjustments.
Bias (in AI)	Systematic errors in AI outputs due to biased training data, leading to unfair or inaccurate predictions.
Black-Box problem	The issue where the internal workings of an AI system, particularly complex models like deep neural networks, are opaque and not easily understood by humans. See also "Algorithmic Transparency" above.
Conditional Text Generation	The process by which a language model generates text based on an input instruction provided in the form of a prompt.
Content Moderation	The process of monitoring and regulating user-generated content on online platforms to ensure compliance with legal, ethical, and community standards.
ControlNet	A technology used in image generation to control and guide the output of AI models, ensuring specific aspects of the image are altered according to predefined rules.
Deepfake	Synthetic media where a person in an existing image or video is replaced with someone else's likeness using AI techniques like GANs. See also "Generative Adversarial Networks (GANs)" below.
Diffusion Models	A type of probabilistic model used to generate data by reversing a diffusion process, often used in generating high-quality images.
Domain Adaptation	A technique in AI where models are adjusted or retrained to improve performance when applied to a new, but related, domain.
Explainability	The capability of AI systems to make their decision-making processes understandable to humans, often linked to transparency and accountability.
Generative Adversarial Networks (GANs)	Machine learning frameworks where two neural networks compete to generate new, realistic data, often used in the generation of deepfakes. See also "Deepfake" above.
Inpainting	An image processing technique where missing or corrupted parts of an image are reconstructed, often used in AI-driven image editing.
Large Language Models (LLMs)	AI models trained on vast amounts of text data to understand, generate, and manipulate human language.
Latent Space	A representation of data in a lower-dimensional space, used in AI models like autoencoders to capture the underlying structure of the data.

Term	Definition
LoRA (Low-Rank Adaptation)	A technique used to fine-tune large AI models efficiently by adjusting a smaller subset of parameters, reducing computational requirements.
Multimodal Models	AI models that process and generate data from multiple types of inputs like text, images, and audio.
Neural Radiance Fields (NeRFs)	Neural networks used to synthesise novel views of complex 3D scenes from sparse input images.
Outpainting	The technique of extending an image beyond its original borders by generating additional content that matches the existing image's style and context.
Pixel Renders	The visual output of an image or scene generated by an AI model, often used in video game testing and content moderation
Procedural Tools	Software tools that use algorithms to automatically generate data or content, commonly used in video game development.
Retrieval Augmented Generation (RAG)	A method combining traditional database queries with AI language models to enhance the accuracy and relevance of generated content.
Self-Orchestrated Workflow	A workflow in which different AI models, such as descriptive, transformative, and generative, work together autonomously to complete tasks.
Tokenization	Short-duration events in audio signals, such as the initial attack of a musical note, which are critical in music generation and processing.
Transients	AI technologies that generate images and video content from textual descriptions, often used in content creation and video production.
Transmodal Analysis	AI approach that generates textual descriptions from multiple data types and processes them for specific analytical objectives.
Uncanny Valley	A technique in AI where models are adjusted or retrained to improve performance when applied to a new, but related, domain.
Vendor Lock-In	Dependency on a vendor for products or services, making it difficult to switch without significant cost or inconvenience.
Vertical and Horizontal Methods (in music)	Techniques in video game music production where music changes are managed linearly (horizontal) or through layered compositions (vertical) to match in-game events.
Vision Language Models (VLMs)	AI models combining visual and linguistic processing capabilities for analysing and generating content based on text and visuals.
Watermarking	Embedding a digital fingerprint in media content to ensure data provenance, integrity, and traceability.

References

1. AI4Media. (2024, July 30). Strategic Research Agenda on AI for the Media industry - AI4Media. https://www.ai4media.eu/strategic-research-agenda-on-ai-for-the-media-industry/
2. https://bellingcat.gitbook.io/toolkit
3. https://weverify.eu/tools
4. https://fullfact.org
5. Wilkinson, C. (2023, February 23). Nigerian fact checkers fight election misinformation with Full Fact's AI Tools. Google Africa Blog. https://blog.google/intl/en-africa/company-news/technology/nigerian-fact-checkers-fight-election-misinformation-with-full-facts-ai-tools/
6. Plaum, A. (2024, February 16). Eight new AI services to support future media verification work. AI4MEDIA Blog. https://www.ai4media.eu/eight-new-ai-services-to-support-future-media-verification-work/
7. https://www.truly.media
8. Anlen, S. & Vázquez Llorente, R. (2024, April 15th). Spotting the deepfakes in this year of elections: how AI detection tools work and where they fail. Reuters Institute Blog. https://reutersinstitute.politics.ox.ac.uk/news/spotting-deepfakes-year-elections-how-ai-detection-tools-work-and-where-they-fail
9. Goyal, S. (2019, June 2nd). GANs — A Brief Introduction to Generative Adversarial Networks. Medium. https://medium.com/analytics-vidhya/gans-a-brief-introduction-to-generative-adversarial-networks-f06216c7200e
10. Erdem, K. (2023, November 9th). Step by Step Visual Introduction to Diffusion Models. Medium. https://medium.com/@kemalpiro/step-by-step-visual-introduction-to-diffusion-models-235942d2f15c
11. Borchani, W. (2023, January 18th). What You Need to Know About Neural Radiance Fields. Medium. https://medium.com/@wiembborchani/what-you-need-to-know-about-neural-radiance-fields-nerf-ba3e5505743f
12. https://openai.com/index/gpt-4
13. https://www.llama.com
14. Abels, G. (2023, May 31). Can ChatGPT fact-check? We tested. Poynter Blog. https://www.poynter.org/fact-checking/2023/chatgpt-ai-replace-fact-checking
15. N.N. (2024, February 6th). Fact checking using artificial intelligence to help journalists. Inria Blog. https://www.inria.fr/en/fact-checking-using-artificial-intelligence-help-journalists
16. https://factly.in
17. Dubbudu, R. & Deshetti, S. (2024, April 4). AI can help fact-checking. We've already seen it. Poynter Blog. https://www.poynter.org/fact-checking/2024/factly-india-artificial-intelligence-fact-checking/
18. Khellaf, L., Baris Schlicht, I., Bayer, J., Bouwmeester, R., Miraß, T. & Wagner, T. (2023, November 10). Spot: A Natural Language Interface for Geospatial Searches in OSM. Proceedings of the OSM Science 2023. https://arxiv.org/pdf/2311.08093
19. https://overpass-turbo.eu
20. https://geospy.ai
21. Kaushik, S. (2023, October 12). Talk to your Database using RAG and LLMs. Medium. https://medium.com/primastat/talk-to-your-database-using-rag-and-llms-42eb852d2a3c
22. Gurrapu, S., Kulkarni, A., Huang, L., Lourentzou, I., Freeman, L. & Batarseh, F.A. (2023, January 21). Rationalization for Explainable NLP: A Survey. Arxiv. https://arxiv.org/abs/2301.08912
23. https://gulbenkian.pt/emifund/projects/the-multilingual-fact-checker-chatbot/
24. Salazar, D. (2023, October 3) How Intel is putting its AI-optimized processors to work detecting deepfakes. Fast Company. https://www.fastcompany.com/90957928/intel-fakecatcher-deepfake-detector
25. https://openai.smapply.org/prog/dalle_detection_classifier_access_program/
26. Bonk, L. (2024, May 7). OpenAI says it can detect images made by its own software... mostly. https://www.engadget.com/openai-says-it-can-detect-images-made-by-its-own-software-mostly-170012976.html
27. https://thehive.ai/demos
28. https://contentauthenticity.org
29. https://www.originproject.info
30. https://liccium.com
31. https://iscc.foundation/iscc
32. https://www.media.mit.edu/projects/detect-fakes/overview
33. Bayer, J., Bouwmeester, R. & Baig, R. (2023, December 6). How do I spot a deepfake? DW. https://www.dw.com/en/fact-check-how-do-i-spot-a-deepfake/a-60029650
34. Chan, K. & Swenson, A. (2024, March 21). One Tech Tip: How to spot AI-generated deepfake images. AP. https://apnews.com/article/one-tech-tip-spotting-deepfakes-ai-8f7403c7e5a738488d74cf2326382d8c
35. https://akademie.dw.com/en/tackling-disinformation-a-learning-guide/s-67264807
36. https://lie-detectors.org
37. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en
38. https://gulbenkian.pt/emifund
39. https://www.poynter.org/ifcn
40. https://www.ebu.ch/files/live/sites/ebu/files/Publications/Reports/open/2024.06.11_Gen-AI-and-PSM_EN.pdf
41. Thomson, T. J., Thomas, R. J., & Matich, P. (2024, April 8). Generative Visual AI in News Organizations: Challenges, opportunities, perceptions, and policies. Digital Journalism, 1–22. https://doi.org/10.1080/21670811.2024.2331769
42. Brecht Declercq. (2024, June 8). Neck Deep in Digital Oil? Public Broadcaster's Archives as AI Training Datasets. FIAT/IFTA. https://fiatifta.org/broadcast-archives-as-datasets/
43. Tushinskiy, S., & Tushinskiy, S. (2024, January 24). How Generative AI Pairs with CTV to Fuel Audience Engagement - MarTech Cube. MarTech Cube - Marketing Technology news and Martech Interviews. https://www.martechcube.com/how-generative-ai-pairs-with-ctv-to-fuel-audience-engagement/
44. Townsend, S. (2024, March 6). The Rise of AI in Film: How AI Script Writing is Changing the Game. Medium. https://medium.com/@channelasaservice/the-rise-of-ai-in-film-how-ai-script-writing-is-changing-the-game-04509a15c9fe

45. Damji, J. S. (2024, April 29). An exploratory Tour of Retrieval Augmented Generation (RAG) paradigm. Medium. https://ai.gopubby.com/an-exploratory-tour-of-retrieval-augmented-generation-rag-paradigm-3940c1947d27

46. Jayaseelan, N. (2024, January 24). How animation industry can be transformed by Generative AI. https://www.e2enetworks.com/blog/how-animation-industry-can-be-transformed-by-generative-ai

47. echo3D. (2023, April 17). 7 Generative AI Tools for 3D Asset Creation - Echo3D - Medium. Medium. https://medium.com/echo3d/7-generative-ai-tools-for-3d-asset-creation-97dd88153b7

48. Mu, Y., Dong, C., Bontcheva, K., & Song, X. (2024, March 26). Large language models offer an alternative to the traditional approach of topic modelling. arXiv.org. https://arxiv.org/abs/2403.16248

49. Springs. (2024, April 8). Prompt Engineering: Examples and best practices - Springs - Medium. Medium. https://medium.com/@springs_apps/prompt-engineering-examples-and-best-practices-82b1da724643

50. Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2024, May 1). A watermark for large language models. arXiv.org. https://arxiv.org/abs/2301.10226

51. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf

52. Kayser-Bril, N. (2020, April 7). Google apologizes after its Vision AI produced racist results. AlgorithmWatch. https://algorithmwatch.org/en/google-vision-racism/

53. informa. (2024). Game Developers Conference. Game Developers Conference. https://gdconf.com/

54. Gamequality.org. (2024). Gamequality.org. Gamequality.org. https://gamequality.org/

55. IQPC. (2024). Game Quality Forum Global. Game Quality Forum. https://www.iqpc.com/events-gamequalityforum

56. Howell, C. (2022). Modern Game Testing: A Pragmatic Guide to Test Planning and Strategy. Modern Game Testing Company.

57. Griffin Gaming Partners and Rendered.vc. (2023). 2023 Game Development Report. https://griffingp.com/wp-content/uploads/2024/02/2023-Game-Development-Report.pdf

58. Agentic. (2024). Agentic Trailer 2024. https://www.agentic.ai/

59. nunu.ai. (2024). nunu.ai. nunu.ai. https://nunu.ai/

60. https://www.asoundeffect.com/stray-game-audio

61. https://unrealcommunity.wiki/metasounds-d660ee

62. https://sonar.es/en/activity/we-are-the-music-makers-ai-and-music-forum

63. Antoine Caillon, and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," Arxiv preprint, 2020. https://arxiv.org/abs/2111.05011

64. https://www.youtube.com/watch?v=dMZs04TzxUI, https://forum.ircam.fr/article/tutoriel-rave-and-nn/, https://forum.ircam.fr/projects/detail/rave/.

65. Théis Bazin, Gaëtan Hadjeres, Philippe Esling and Mikhail Malt. "Spectrogram Inpainting for Interactive Generation of Instrument Sounds," Joint Conference on AI Music Creativity, 2020. https://boblsturm.github.io/aimusic2020/papers/CSMC_MuMe_2020_paper_49.pdf

66. https://www.youtube.com/watch?v=SzlxPTIcxnY, https://sonycslparis.github.io/interactive-spectrogram-inpainting/

67. https://cslmusicteam.sony.fr/prototypes/bassnet/, https://www.youtube.com/watch?v=Nugl8Xqs3d4

68. Maarten Grachten, Stefan Lattner, and Emmanuel Deruty. "BassNet: A Variational Gated Autoencoder for Conditional Generation of Bass Guitar Tracks with Learned Interactive Control," Applied Sciences 10(18), 2020. https://doi.org/10.3390/app10186627

69. Marco Pasini, Maarten Grachten, and Stefan Lattner, "Bass accompaniment generation via latent diffusion," IEEE ICASSP, 2024. https://doi.org/10.1109/ICASSP48485.2024.10446400

70. Andrea Agostinelli, et al. "MusicLM: Generating music from text," Arxiv preprint, 2023. https://arxiv.org/abs/2301.11325

71. https://google-research.github.io/seanet/musiclm/examples/

72. Jade Copet, et al. "Simple and controllable music generation," in Advances in Neural Information Processing Systems, 2024. https://arxiv.org/abs/2306.05284

73. https://suno.com/

74. https://www.udio.com/

75. https://www.beatoven.ai/

76. https://www.youtube.com/watch?v=dMDwtlV_Dz8, https://musescore.com/egriff123/scores/7257084

77. https://www.youtube.com/watch?v=zkZGFNCS6iQ

78. https://www.youtube.com/watch?v=jpyVfd8TyoI

79. https://www.youtube.com/watch?v=eqJFVg05b8Q

80. https://www.youtube.com/watch?v=3CSO88UekXE

81. "Also sprach Zarathustra" by Richard Strauss, "The Blue Danube" by Johann Strauss II, "Lux Aeterna" and "Atmosphères" by György Ligeti.

82. Kalingeri, V., & Grandhe, S. (2016). Music Generation with Deep Learning. ArXiv.org. https://arxiv.org/abs/1612.04928

83. van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. ArXiv.org. https://arxiv.org/abs/1609.03499

84. Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., & Yang, Y.-H. (2018). MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1709.06298

85. Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., & Défossez, A. (2023, June 8). Simple and Controllable Music Generation. ArXiv.org. https://doi.org/10.48550/arXiv.2306.05284

86. Google Deepmind. (2023, November 16). Transforming the future of music creation. Google DeepMind. https://deepmind.google/discover/blog/transforming-the-future-of-music-creation/

87. Generating audio for video. (2024, June 17). Google DeepMind. https://deepmind.google/discover/blog/generating-audio-for-video/

88. https://ai.meta.com/resources/models-and-libraries/audiocraft/

89. https://www.microsoft.com/en-us/research/project/ai-music/

90. https://openai.com/index/jukebox/

91. Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

AI4media

Follow us @ai4mediaproject

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911
The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.

info@ai4media.eu www.ai4media.eu