THE REVOLUTION IN VISUAL CREATION:
Generative Artificial Intelligence

MIGUEL CASAS ARIAS[1], ALEJANDRO PRIEGO DÍAZ[1], MARÍA LARA-MARTÍNEZ [1]
[1]Complutense University of Madrid, Spain

KEYWORDS

Artificial intelligence
Photography
Midjourney
Visual
Social Networks

ABSTRACT

The integration of artificial intelligence (AI) in audiovisual creation is redefining the boundaries between human creativity and technological potential. Its use is widespread in social networks.
This research will review the technical background and aims to analyse the application of AI in the different stages of visual production. It will study whether communication professionals can take advantage of their knowledge to get more out of these tools. The conclusions determine that AI is involved in the emergence of new forms of artistic and communicative expression.

1. Introduction

As the volume and speed of big data continue to expand, it has been assumed for some years now that artificial intelligence (AI) will play an increasingly important role in media-related contexts (Forrester, 2017). Multimedia content producers are utilising AI to enhance efficiency and accelerate the performance of multimedia production (Jayanthiladevi et al., 2020).

However, this presents a significant challenge, as the most prevalent method of human-machine interaction in AI technology is natural language processing (NLP). This is particularly problematic when dealing with images and descriptions of real-world language, which are inherently noisy and contain linguistic variability and inconsistencies in conveying complex concepts (Zhang and Peng, 2018).

The combination of audiovisual technology and AI has led to the development of systems for generating images from text, which have become popular tools for the creation of digital images and works of art (Crowson et al., 2023). In the present era, audiovisual and artistic professionals, as well as social media users lacking specialized training, are able to generate their content using these applications. Moreover, the capacity to create content is now within the reach of any Internet user, who can simply enter text in natural language to initiate the process. In the scientific literature analysed, it is referred to by different names, but it is commonly known by the term used for the messages that are introduced into the different systems to generate instructions to the AI: prompts.

These content creation systems combine generative algorithms with an intuitive interface for entering natural language prompts and receiving generated images that convey the concepts and features expressed in the text. The accessibility of these systems has been significantly enhanced in recent times, thereby increasing public awareness of their capacity to generate high-resolution images. The results are both surprising and original, providing aesthetic styles that match textual indications with a high degree of accuracy (Brisco et al., 2023). This study therefore raises the question of the extent to which text-to-image generative AI could influence or contribute to the design and creation of professional visual content, in particular still images. Furthermore, AI is being integrated into all aspects of audiovisual production, including film, television, video, social media, and broadcasting. This research examines the trajectory of creativity, productivity, and the quality of results in photography as they relate to the application of creativity in photography.

There are significant ethical concerns when incorporating AI into art. The question of authorship and copyright remains unresolved, while the prospect of job destruction is a recurring theme. When an AI generates a piece of art, it is unclear who may be considered the artist: the individual who created the algorithm, the individual who implemented it, or the AI system itself. Moreover, in numerous fields, including the arts, the potential for bias in AI algorithms is an ethical concern. The potential exists for the art produced to perpetuate stereotypes or reflect biased viewpoints if the data used to train these systems contains bias.

1.1. Objectives of the research

The objective of this research is to achieve two specific goals. The initial objective is to situate AI in the context of audiovisual production in the media and social networks. This will entail an analysis of applications that utilise AI solutions and their functionality. Secondly, the efficacy of the most prevalent text-based image generation tools will be evaluated. This study will examine the challenges and emerging needs related to the use of these software applications, focusing on the nomenclature and tools used, and the need to obtain professional results compatible with human creative decisions.

The objective of this article is to gain an understanding of how AI is changing the process of content creation in the visual realm and how this is affecting the emergence of new forms of artistic and communicative expression.

1.2. Justification

A considerable corpus of literature exists on the applications of AI, yet there is a paucity of scholarship that has focused on the use of AI in audiovisual media. Additionally, there has been no study that has focused on the shortcomings of AI as a medium for the generation of visual content within parameters that reach a professional level. The pace of technological advancement is such that, on a weekly basis, large companies specialising in this field implement increasingly sophisticated technical innovations. This has led to a point in time where it is necessary to clarify the scope and position of AI applications in the visual environment.

2. State of Play

2.1. Contextualising AI in the Audiovisual Environment

Since the late 1950s, several novel artistic practices linked to technology have emerged. A clear and synthetic description of these practices has been provided by Margaret A. Boden and Ernest A. Edmonds in their article "What is Generative Art" (Boden and Edmonds). In their article "What is Generative Art" (Boden and Edmonds, 2009), the authors frame the current state of AI technology development within the context of content generation, categorising it as follows:

- Computer art: Art in which computers are involved in the production process.
- Electronic art: This expansive concept encompasses any artistic work whose production involves electrical engineering and/or electronic technology.
- Generative art: Artworks that are at least partially generated by a process that is not directly under the artist's control.

These technologies have frequently been discussed in terms of their aesthetic implications in relation to art theory. Specifically, and as a consequence of their exponential growth, social networks have witnessed the most significant surge in the utilisation of novel technologies for creative expression, particularly in the wake of the pandemic, far exceeding any other field (Arana Arrieta et al., 2020).

The current growth of AI began with the development of GPT (Generative Pre-trained Transformer) Chat in 2022, a technology that is integrated into generative AI. The field of study deals with the creation of forms of communication between humans and machines using everyday language, both spoken and written. Although the first contemporary representation of the GPT chat model was first observed in 2018, subsequent developments have led to a notable increase in both the quality and complexity of the technology.

Images generated by AI in previous years exhibited a relatively low resolution, resulting in images that were dark and lacked detail (Zhang and Peng, 2018). It is our contention that the present evolution will represent a significant milestone in the field of technology. It is becoming increasingly evident that AI is being employed by filmmakers in an increasing number of cases (Momot, 2022).

2.2. Automatic Generation of Audiovisual Content with AI

These programmes are capable of generating entirely novel images from a text with a description based on a language of words. The information in the text is processed to generate an image that attempts to match the description. The advent of artificial intelligence (AI) and the subsequent development of deep learning techniques, coupled with the availability of vast datasets and the requisite computing power, has made it possible for computers to generate images from textual descriptions and, conversely, to generate textual descriptions for images (Brisco et al., 2023; Zhang and Peng, 2018).

In the field of audiovisual production, AI has made significant advances in automating tasks that previously required human intervention, particularly in the area of automatic content generation. In their paper, "Image Style Transfer Using Convolutional Neural Networks," Gatys and other researchers discuss the development of algorithms capable of learning artistic visual representations and applying specific styles to images. They demonstrate the power of AI in visual creation (Gatys et al., 2016). Although generative AI had already demonstrated its efficiency in content generation, its implementation was only permitted until 2020 (Son et al., 2019). The application of generative AI in industries such as film, digital platforms and broadcasting is significantly limited. The current potential of generative AI technology is most evident in the domain of digital images and video (Lee, 2023).

The launch of the Midjourney community represented a pivotal moment in this regard. Members are able to view images created by other members, along with the prompts used. The prompts are therefore shared within the community, thus becoming a social learning resource, which enables the creativity of what was previously a single creator practice to be unlocked (Oppenlaender, 2022). In this manner, a multitude of prompts for modifying the style and quality of images have been identified and are being consistently implemented within the Midjourney community (Oppenlaender, 2023).

2.2.1. Text to Image Generation

In 2021, considerable progress was made in the field of text-to-image generation for the synthesis of digital visual content (Cobb, 2023). These systems have already reached a stage of development that allows them to generate aesthetically pleasing images, especially when prompts are used to alter the style or improve the quality of the image (Brisco et al., 2023). Nevertheless, in numerous instances, additional processing steps were required, such as eye correction in human subjects (Nightingale and Farid, 2022).

In most cases, particularly on social media, where images from non-professional users are the predominant source, text-to-picture art is driven by exploration. Users test their ideas or models by providing textual prompts to ascertain which elements are effective and which are not. This method could also be extended to professionals. Consequently, the process is one of gradual evolution, whereby ideas are connected between subsequent prompts. It is not uncommon for a subjectively satisfactory outcome to require multiple interactions (Oppenlaender, 2022). Furthermore, the process is influenced by the AI's own interpretation and the ambiguity of words in representing ideas (Zhang and Peng, 2018). At this juncture, it is the user's prerogative to permit the discourse with the AI to diverge in unforeseen directions. Similarly, practitioners may also follow the opportunities presented during the development of their creations by chance.

Consequently, contemporary users of non-professional text-to-picture applications are compelled to meticulously select a representative sample of their most accomplished work, while discarding (or electing not to publish) other images. The process of creating text-to-picture art is analogous to that of a photographer, who filters and eliminates a selection of images from a larger collection in order to rapidly explore a greater number of alternatives. This represents a significant competitive advantage. The Midjourney community has incorporated a selection process into the platform, whereby only images that have been enhanced by the user are displayed by default. This practice allows for the application of holistic filtering metrics to assess the creativity of the prompter (Oppenlaender, 2022).

In their article "Design Guidelines for Prompt Engineering Text-to-Image Generative Models," Liu and Chilton (2022) investigate the generative capabilities of text-to-image generation. In this article, the authors present an analysis of the results of prompts that have gained popularity among experts. The objective is to establish design standards that facilitate superior results through the use of this technology.

2.2.2. Text to Video Generation

This branch of generative AI is currently undergoing a phase of experimentation and rapid growth. It is anticipated that this period of experimentation will be relatively brief, during which time the technology will be refined and will begin to demonstrate optimal results on a large scale.

The generation of video requires a significantly greater degree of computing power and memory than the generation of still images. This is due to the necessity of creating multiple images (frames) in order to recreate the movement of a digital image. Furthermore, it is essential to maintain aesthetic and content coherence and fluidity of movement throughout the sequence. This process entails not only the generation of multiple images but also the comprehension and implementation of cinematic principles, such as continuity and visual narrative (Fu et al., 2023).

As of February 2024, the image quality of AI-generated videos has reached a remarkable level of achievement. However, it has not yet reached the level of realism that can be achieved with still images.

2.3. Post-Production Tools

The development of artificial intelligence (AI) applications in the field of audiovisual post-production represents a notable advance with each innovation. The distinctive feature of these applications is their capacity to specialise in different utilities within the creative process. In recent months, new tools and more sophisticated versions have been introduced on a weekly basis. In the domain of image and video capture, algorithms can automate processes such as focusing, lighting, and composition. Furthermore, in the domain of editing, AI facilitates the identification of key elements, enhances image quality, and even proffers creative options for visual storytelling. The post-production possibilities include tools that enable the alteration of faces, the creation of original art, the removal of backgrounds from images, the

improvement of image quality, the inclusion of elements in videos, the adaptation of a speaker's lips to another language, the generation of songs and sound effects, the generation of voices, the dubbing of voices into other languages, the cloning of voices, the fixing of out-of-focus videos, the animation of 3D characters and objects, the creation of interactive content for video games and VR, the analysis and classification of visual and auditory content, the generation of subtitles, the translation of audio in real time, and the facial and object recognition in videos.

One of the most prevalent applications in the domain of post-production is prompted-based image editing. Although human instructions are sometimes too brief for current methods to execute them properly (Fu et al., 2023), it is already possible to automate the task of content management, audio/video synchronisation, multimedia operations and reduce human errors (Jayanthiladevi et al., 2020).

One of the most significant tools is the capacity to fill in internal parts of images, which is known as image inpainting (Zhang et al., 2020). The term 'inpainting' is used to describe a specific technique for reconstructing missing or damaged parts of images or videos. This is achieved with artificial intelligence algorithms, which are employed to fill in the spaces in a way that is consistent with the rest of the image (Elharrouss et al., 2020). These innovative creative techniques offer a level of interactivity that extends beyond the mere generation of static images from text input prompts (Oppenlaender, 2023). This technology represents a significant advancement in the field of photographic and video post-production.

3. Design and Method

3.1. Hypotheses

Those engaged in content generation and decision-making, whether they are network users or media professionals, must not only generate creative ideas but also assess their quality in order to produce the desired innovation and impact. In this domain, both the networks and the professional world are sustained by a vast array of photographic material. To illustrate, one of the most widely utilized networks for content generation is Instagram, which specializes in photography and video.

In light of the generative capacity of AIs, it is pertinent to inquire whether text-based image generation tools are capable of producing photographs that resemble those taken by a professional photographer in a traditional manner and can be used on networks or in professional publications without users being able to distinguish them from real photographs.

In addition, it is necessary to determine which prompts are more useful for the generation of photographic images. Are prompts that are merely artistic and descriptive more effective, or are prompts that contain professional photographic parameters more beneficial?

3.2. Methodology

Firstly, a documentary review of the existing scientific literature has been carried out in order to provide an objective analysis of the role of AI in the audiovisual industry. A comprehensive examination of the available data was conducted, with a particular focus on one of the most rapidly expanding applications of AI: the generation of content from text to image, text to video, and the specific applications associated with this process.

A review of the literature in Scopus and Web of Science revealed a paucity of publications on the subject. However, Google Scholar contains a considerable number of publications, although the majority of these are oriented towards particular analyses of the technological novelties of developers without greater research pretensions.

Conversely, an experimental methodology is proposed utilising the leading image generation tools currently available on the market. The following three applications were selected for further analysis: Midjourney, Stable Diffusion and Dall-e 3. To arrive at this result, a preliminary investigation was conducted via Google Trends on the three most searched applications on Google among the 20 most popular ones over the past year.

Figure 1. Google Trends Comparison Chart
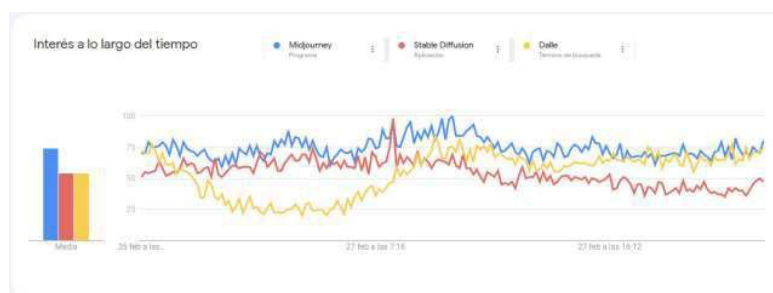
Source: Google Trends, 2024.

Building on this foundation, a subsequent analysis will be conducted to assess the realism of the images created by the three selected applications. This analysis will be based on several parameters.

Subsequently, identical prompts will be provided to these applications, accompanied by instructions in technical terminology, with the objective of evaluating and comparing their ability to respond to specific professional commands. This methodology will permit an impartial comparison of the manner in which each system processes and handles the identical prompts, thereby providing a robust foundation for the comprehension of the strengths and limitations of each tool in the context of professional image generation. This, in turn, will permit the determination of whether the result can be considered professional and, most importantly, whether it is in accordance with the preliminary ideas of the creator of the idea.

The experiment will comprise a comparison of traditional methods of photographic creation and the generation of similar work using generative AI. To this end, we will commence with the work of two professional photographers, renowned for their contributions to the field, and endeavour to emulate their work through the use of deep generative models. In doing so, we will introduce a series of prompts of both descriptive and technical levels, derived from the photographers' own photographs.

The results will be evaluated through a quantitative and qualitative analysis of first-hand experience and the results obtained using both techniques. The results obtained through the use of AI will be compared with the actual photographs taken as a working basis.

To ascertain the feasibility of this approach, the initial phase of the experiment involved providing a basic prompt for the generation of a photograph in three text-to-image generative AIs. In this instance, the following textual description has been employed to generate the requisite photographs. The prompt was "a photograph of Concha Beach in Donosti".

Upon examination of the generated images, it becomes evident that there are significant discrepancies between the outputs of the various models. In the case of DALL-E 3 (Figure 2), the result can be described as an illustration. It can be observed that the image resembles a painting, with no elements that are faithful to reality. Instead, it is perceived as an artistic representation of the selected place. In Stable Diffusion (Fig. 3), although the result is more realistic, the lack of detail precludes the conclusion that it is a professional photographic representation. This is due to the indistinctness of both the people and the facades of the buildings depicted.

While Midjourney has generated images that are consistent with the proposed prompt, the result is four images that simulate a photograph of La Concha Beach (Fig. 4). The sand, water, people, landforms and buildings are clearly distinguishable and, moreover, resemble to a high degree what is perceived in a normal photographic image. Consequently, the remainder of the experiment will focus on the utilisation of the Midjourney model.

Figure 2. Image of La Concha beach created with DALL E-3

Source: DALL E-3, 2024.

Figure 3. Image of La Concha beach created with Stable Diffusion
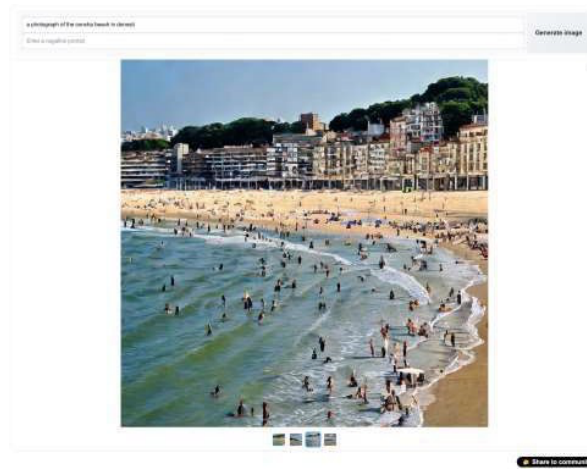
Source: Stable Diffusion, 2024.

Figure 4. Images of La Concha Beach created with Midjourney.

Source: Midjourney, 2024.

Furthermore, Midjourney is straightforward to use, thereby reducing the time required to become proficient in its use to a significantly shorter period than with other software. This is due to the necessity of generating a large number of images before studying certain prompts (Liu and Chilton, 2022).

The subsequent stage entails the generation of images that resemble photographs through a series of prompts that emulate the work of professional photographers with a highly emblematic oeuvre. In order to achieve this objective, it has been determined that the works of Martin Parr and Ansel Adams will serve as the foundation for this endeavour.

Martin Parr is a renowned British photographer and member of the Magnum agency. His work is primarily documentary photography in a genre style, characterised by a distinctive sense of humour. The experiment is based on two photographs from Parr's Life's a Beach series from the mid-1990s (Fig. 5 and 6).

Figure 5. Photograph by Martin Parr from the series *Life's a beach.*

Source: www.martinparr.com, 1997.

Figure 6. Photograph by Martin Parr from the series *Life's a beach.*

Source: www.martinparr.com, 1997.

Ansel Adams was an American photographer renowned for his monochromatic images of natural landscapes within the United States. The following works were employed in the experiment: The images were captured in Glacier National Park, Montana (Fig. 7) and Long's Peak, Rocky Mountain National Park, Colorado (Fig. 8).

Figure 7. Photograph by Ansel Adams *in Glacier National Park*

Source: archives.com.gov, 1942.

Figure 8. Photograph of Ansel Adams *Long's Peak*

Source: archives.com.gov, 1942.

Three distinct experiments are conducted, each employing a distinct textual description to generate the requisite images. The objective of each experiment is to generate an image that is as close as possible to the work of both authors.

4. Fieldwork and Data Analysis

In order to conduct the experiment, an image generation session is carried out using English prompts in order to achieve the desired image in Midjourney. A diverse range of prompts are trialled, with the aim of identifying the impact of varying descriptions, styles and word counts on the resulting images.

These are subsequently saved in a high-resolution PNG format for further analysis. A total of 125 original images were generated based on the work of the two reference authors between 20 and 26 February 2024.

The generated images are then subjected to qualitative analysis based on several criteria. These include an assessment of the degree of realism of the elements generated, the extent of similarity with the reference works, and the degree of correspondence with the selected prompt.

In order to illustrate the current capabilities of generative text-to-image AI, particularly the Midjourney model, a selection of representative examples has been chosen to demonstrate the potential for reproducing photographic images.

5. Results

At the outset of the experiment, the use of prompts is confined to a relatively simple textual description. Midjourney is then tasked with generating a photograph in the style of each of the photographers mentioned, along with another photograph specifying the content of the photograph.

Initially, the AI is instructed to generate images in the style of Martin Parr, which is defined as "a photography in Martin Parr's style." Midjourney returns four images that are quite realistic and could be considered photographic images (Fig. 9). These images exhibit a distinct style that is closely aligned with Martin Parr's work, featuring characteristic colours and lighting.

Figure 9. Images generated in Midjourney in the style of Martin Parr.

Source: Midjourney, 2024.

Regarding the work of Ansel Adams, a comparable prompt is generated to produce the images, in this instance Midjourney is requested to generate a photograph in Ansel Adams' style (i.e. a photograph in Ansel Adams' style) (Fig. 10). The result of this process is the generation of four images of landscapes that are highly reminiscent of those captured by Ansel Adams. These images feature snow-capped mountains, expansive pine trees, and aquatic elements such as rivers and lakes. Additionally, reflections of other parts of the image can be observed.

Figure 10. Images generated in Midjourney in the style of Ansel Adams.
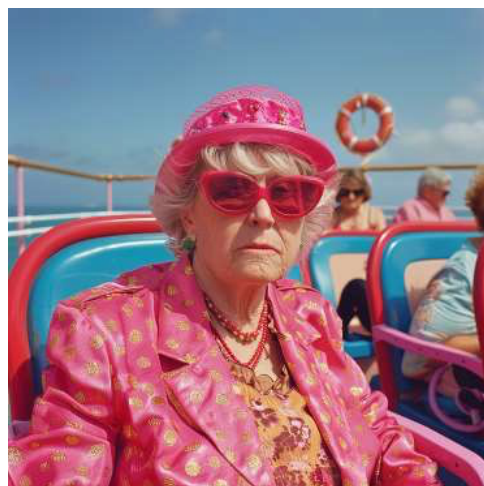

Source: Midjourney, 2024.

In a subsequent stage, one of the options available from Midjourney is employed to incorporate a more technical prompt into the images. This option permits the modification of some of the four generated images by the inclusion of new texts. This is accomplished with the images generated for both authors. The result of the AI is the generation of four additional images based on the selected image from the set of previously generated images.

In the case of Martin Parr, the second image has been selected, namely that of a woman with pink glasses and a hat on a boat (Fig. 11). In the case of Ansel Adams, the image of snow-capped mountains with a lake in the foreground and the reflection of the mountains in it (Fig. 12) has been selected.

Figure 11. Image generated in Midjourney in the style of Martin Parr.

Source: Midjourney, 2024.

Figure 12. Image generated in Midjourney in the style of Ansel Adams.

Source: Midjourney, 2024.

The image of the woman is to be used as a source for the generation of other images, which are to be captured using an 18 mm lens and a shallower depth of field ("a photography in Martin Parr style with a lens of 18 mm and less depth of field"). The result is four images (Fig. 13) that are similar to the previous one in which a woman with glasses and a pink hat can be seen in the foreground. The backgrounds are completely out of focus. It can also be seen that none of the images corresponds to what would be a photograph taken with an 18 mm lens. Conversely, the textures of the garments and the skin are similar to those of the real thing. With the exception of the detail of a coffee cup of an unusual shape on the edge of a table, there are no major errors in the images.

Figure 13. Images generated in Midjourney in the style of Martin Parr.

Source: Midjourney, 2024.

In the case of the Ansel Adams-style image, Midjourney is tasked with generating an image using an 8mm lens and a shallower depth of field, which can be described as a "photography in Ansel Adams style with an 8mm lens and less depth of field." The resulting images (Fig. 14) are four images that are similar to the previous one. While it can be observed that the optics could be an angular lens, it is clear that the 8mm lens has not been reached, and the depth of field is not shallow.

Figure 14. Images generated in Midjourney in the style of Ansel Adams.

Source: Midjourney, 2024.

In the second phase of the experiment, we proceed as follows: we take as a basis the works of both authors selected at the outset of the study and proceed to describe these photographs in order to create a purely descriptive prompt based on them. This is to enable Midjourney to generate new images.

To generate an image similar to those in Martin Parr's Life's is Beach series (Fig. 5 and 6), the following prompt is introduced: "A photograph of people seated on the sand with sun loungers and umbrellas in the sun. Some sun loungers are unoccupied, while others are in use by individuals. Additionally, other individuals are reclining on towels. All sun loungers and umbrellas are of a bright colour. It should be noted that not all of the beach is occupied. The image depicts the sea and a number of individuals engaged in aquatic activities. The photograph was taken from a height of 1.75 metres with an 18mm lens".

Figure 15. Images generated in Midjourney describing the figure. 6

Source: Midjourney, 2024.

The generated images (Fig. 15) exhibit a high degree of realism. All of the parameters used for their creation have been fulfilled, and the sand and water textures are accurate. The umbrellas, deckchairs, and other beach elements are depicted with a high degree of verisimilitude. Nevertheless, it is evident that certain individuals depicted in the images lack sufficient representation. These include implausible body positions, incomplete pieces of clothing, poorly defined factions, and errors.

In order to create an image based on Ansel Adams' photographs (Fig. 7 and 8), the following prompt was employed: "The photograph in question is a black and white image of a view from above of a plateau with some pine trees, a snow-covered mountain in the background in Long Peak, at Rocky Mountain National Park. The photograph was captured with a 35mm objective lens".

Figure 16. Images generated in Midjourney describing the figure. 7

Source: Midjourney, 2024.

As evidenced by the four exemplary images presented, the requisite standards set forth by Midjourney have been met.

The subsequent stage is to incorporate images in the authors' style into the aforementioned prompts. To achieve this, one of the generated images is selected and employed to create the new images to be incorporated into the text (Fig. 17 and 18).

Figure 17. Image generated in Midjourney.

Source: Midjourney, 2024.

Figure 18. Image generated in Midjourney.

Source: Midjourney, 2024.

Including the description that the images are in the style of Martin Parr has the effect of producing four photographs (Fig. 15) with the same positive features in the representation that the AI has relied on for its generation. These include the inanimate elements and water, which are well imitated. However, the same shortcomings as before are evident, namely that human bodies are not quite well defined, with some parts cut off or deformed.

Figure 19. Images generated in Midjourney depicting Figure 6.

Source: Midjourney, 2024.

With regard to the images created in the style of Ansel Adams (Fig. 20), it can be stated that they adhere to the same literal interpretation of the description provided for their predecessor and are able to reproduce the fundamental elements of a photograph by the author without any difficulty.

Figure 20. Images generated in Midjourney depicting Figure 8.

Source: Midjourney, 2024.

The table presented below is integrated into our study in order to provide a detailed and quantitative analysis of several crucial parameters in the evaluation of AI-generated images. Each category is evaluated on a spectrum between 1 and 5, thus providing an objective comparative metric. The table serves to facilitate understanding and establish a comparative framework for future research.

Table 1. Analysis of Realism and Coherence of Generated Images

	Suitability/approximation to a picture	Coherence of elements
Landscapes 1ˢᵗ flat	12345	12345
Landscapes 2nd plane	12345	12345
Faces 1ᵉʳ flat	12345	12345
Faces 2ᵒ flat	12345	12345
Extremities 1ᵉʳ flat	12345	12345
Extremities 2ᵒ flat	12345	12345
Objects 1ᵉʳ flat	12345	12345
Objects 2ᵒ flat	12345	12345
Textures	12345	12345
Simulation of style	12345	12345
Lighting	12345	12345
Depth of field	12345	12345
Lens selection	12345	--------
Eye Level	12345	--------

Source: Own elaboration, 2024.

6. Discussion and Conclusions

6.1. Discussion

The capacity to compose efficacious prompts for interaction with content creation tools is becoming a pivotal area of expertise. As Oppenlaender (2023) posits, the efficacy of these prompts is contingent not only on the technical expertise of neural networks, but also on the individual's experience and familiarity with various request modifiers (prompts). It has been demonstrated that a comprehensive comprehension of the functioning of these networks and the means of modifying their inputs to achieve the desired outcomes is a prerequisite for optimising their potential. We concur that rapid message engineering is a process of learning and skill acquisition, given the lack of immediate clarity regarding the optimal methodology for writing prompts and the identification of effective keywords as modifiers (Rogers et al., 2020). It is our contention that the interpretation and correct use of prompts necessitates experience. We question whether the aforementioned skills encompass the more technical and professional nomenclatures, or whether prompts should be limited to more general descriptions.

Among the challenges identified through our analysis, it was found that unintended restricted data input limits the exploitation of AI. These data constraints present a significant technical obstacle to achieving the desired results. Furthermore, it is important to acknowledge that the design and programming of AIs is undertaken by human developers, who may introduce biases into the process. It is acknowledged that bias is a human problem, not a machine problem. Consequently, the approach to addressing it cannot rely exclusively on technical solutions (Schetinger et al., 2023). Nevertheless, the experiment has demonstrated that the technical limitations remain significant, and that not all hallucinations generated by AI can be resolved by introducing new prompts designed to modify the results.

When designing audiovisual content through AI, with a criterion aimed at a specific objective, whether personal or commissioned, but not random, it is important to have the concepts of the result we intend to achieve well defined, with specific characteristics of structure, style and message. These characteristics are essential for achieving the functional objectives of the content, such as conveying a specific message, telling a story or generating an emotional response in the audience (Brisco et al., 2023). Nevertheless, there are currently certain limitations to be considered. It is evident that these systems lack the capacity to fully comprehend the structural and behavioural attributes delineated in the text. It has been observed that, in the absence of a specific style request, the generated scenes may appear surreal and dreamlike, deviating from the conventional norms of real-world or traditional

audiovisual content. In practice, this implies that if a filmmaker or content creator attempts to utilise text-to-picture AI, they may initially arrive at results that appear unrealistic or artistically abstract. While this may be beneficial for creative projects that seek a novel, unique or surreal visual style, it is not suitable for projects that require accurate and realistic representations based on detailed concepts. The advent of AI has created a situation in which the prevalence of amateurism, particularly on social media, has made it easier for individuals lacking artistic or technical expertise to develop their ideas. AI-based systems represent an extension of our human cognitive abilities, with the potential to enhance and support our creativity. However, the capacity to generate creative ideas is not a sufficient condition for innovation (López et al., 2019). We concur with López that in order to create engaging content through AI, content creators and decision-makers, whether they are network users or media professionals, must not only generate and conclude on creative ideas but also deliver quality results in order for innovation and the desired impact to occur.

The advent of text-to-image technology has led to a predictable and growing interaction with these systems. As this relationship continues, our communication skills with AI will evolve. This novel relationship with computing may have an impact on our behaviour, language, knowledge and skills. As with the evolution of the World Wide Web, the interaction with AI in natural language will shape the future digital society, the way we interact with computers and the way we work (Oppenlaender, 2022).

Nevertheless, it is our contention that the incorporation of novel technologies into the framework of a professional team necessitates a minimum level of awareness on the part of content creators with regard to the critical aspects that these technologies can bring to the table. The necessity for designers to possess a wide range of specific skills in order to keep up with the technological dynamism of our society presents a challenge in the creation of products that are well aligned with user needs, given the introduction of new tools and competences (Figoli et al., 2022).

6.2. Conclusions

This research has demonstrated that AI has already been implemented as a creative and productivity assistant in a manner that aligns with the current reality. The benefits of AI will allow it to become increasingly prevalent, facilitating content personalisation. In the future, it will become essential to optimise operational efficiency during production.

The fusion of human and machine creativity that is enabled by generative art offers a novel perspective on artistic expression. The adept and proficient manner in which prompts are employed serves to reinforce creativity. The accessibility of these tools has enabled a new generation of self-taught, unskilled creators to flourish.

At present, generative AI can produce content of varying degrees of professionalism, contingent on the resources and time invested. With regard to the generation of photographic visual representations that are faithful to reality through the use of a descriptive text, the images produced by the Midjourney model exhibited a high degree of similarity with respect to the work of the proposed authors. Furthermore, the model demonstrated remarkable precision in interpreting the proposed text.

Nevertheless, focusing the analysis on the interpretation of certain more technical concepts, such as depth of field or a particular optic, the results are not entirely satisfactory. This is particularly evident in the representation of complex human elements in large quantities and without being part of the main plane of the photograph.

At present, AI models lack the flexibility to make fine adjustments or specific modifications that a practitioner may require, which limits their usefulness in specialised production environments. Inconsistency in AI-generated results and lack of predictability can be problematic in a professional environment where consistency and reliability are required.

It is too early to assess which of these new tools will become indispensable, or whether they will evolve into another technology and go out of fashion once the novelty wears off. It remains to be seen whether these collaborative projects will ultimately result in works that combine human creativity with the analytical efficiency of AI, or whether a gimmicky, low-budget audiovisual future will prevail, with AI assuming most processes.

References

Adams, A. (1942). National Archives. https://www.archives.gov/espanol/ansel-adams

Arana Arrieta, E., Mimenza Castillo, L. & Narbaiza Amillategi, B. (2020). Pandemia, consumo audiovisual y tendencias de futuro en comunicación. *Revista de Comunicación y Salud*, *10*(2), 149–183. https://doi.org/10.35669/rcys.2020.10(2).149-183

Boden, M. A. & Edmonds, E. A. (2009). What is generative art? *Digital Creativity*, *20*(1–2), 21–46. https://doi.org/10.1080/14626260902867915

Brisco, R., Hay, L. & Dhami, S. (2023). Exploring the role of text-to-image AI in concept generation. *Proceedings of the Design Society*, *3*, 1835–1844. https://doi.org/10.1017/pds.2023.184

Chen, L., Wang, P., Dong, H., Shi, F., Han, J., Guo, Y., Childs, P. R. N., Xiao, J. & Wu, C. (2019). An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation, 61,* 10-22. https://doi.org/10.1016/j.jvcir.2019.02.009

Cobb, P. J. (2023). Large Language Models and Generative AI, Oh My! *Advances in Archaeological Practice, 11,* 363–369). *Cambridge University Press*. https://doi.org/10.1017/aap.2023.20

Crowson, K., Biderman, S., Kornis, D. & Stander, D. (2023). VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. *Arxiv. Cornell University.* https://doi.org/10.48550/arXiv.2204.08583

Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Akbari, Y. (2020). Image Inpainting: A Review. *Neural Processing Letters, 51,* 2007–2028. Springer. https://doi.org/10.1007/s11063-019-10163-0

Evans, Z., Carr, C., Taylor, J., Hawley, S. H., & Pons, J. (7 febrero 2024). Fast Timing-Conditioned Latent Audio Diffusion. *Arxiv. Cornell University.* https://doi.org/10.48550/arXiv.2402.04825

Figoli, F. A., Mattioli, F., & Rampino, L. (2022). AI in the design process: training the human-AI collaboration. *Proceedings of the 24th International Conference on Engineering and Product Design Education 2022*. *The design society.* https://doi.org/10.35199/EPDE.2022.61

Forrester Consulting (5 octubre 2017). *The Machine on your Team: New study shows how marketers are adapting in the Age of AI*. https://www.prnewswire.com/news-releases/the-machine-on-your-team-new-study-shows-how-marketers-are-adapting-in-the-age-of-ai-300531385.html

Fu, T.-J., Hu, W., Du, X., Wang, W. Y., Yang, Y., & Gan, Z. (2023). Guiding Instruction-based Image Editing via Multimodal Large Language Models. *Arxiv. Cornell University.* https://doi.org/10.48550/arXiv.2309.17102

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). *Image Style Transfer Using Convolutional Neural Networks*. Computer Vision Foundation. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf

Jayanthiladevi, A., Raj, A. G., Narmadha, R., Chandran, S., Shaju, S., & Krishna Prasad, K. (2020). AI in Video Analysis, Production and Streaming Delivery. *Journal of Physics: Conference Series*, *1712*(1). https://doi.org/10.1088/1742-6596/1712/1/012014

Lee, S. (2023). Transforming Text into Video: A Proposed Methodology for Video Production Using the VQGAN-CLIP Image Generative AI Model. *International Journal of Advanced Culture Technology*, *11*(3), 225–230. https://doi.org/10.17703/IJACT.2023.11.3.225

Liu, V., & Chilton, L. B. (2022, April 29). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3491102.3501825

López, C. E., Miller, S. R., & Tucker, C. S. (2019). Exploring biases between human and machine generated designs. *Journal of Mechanical Design, Transactions of the ASME*, *141*(2). https://doi.org/10.1115/1.4041857

Mirowski, P. W., Mathewson, K. W., Pittman, J., & Evans, R. (2023). Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. *CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3544548.3581225

Molina-Siles, P., & Giménez Ribera, M. (2023). Inteligencia artificial y creatividad para la generación de imágenes arquitectónicas a partir de descripciones textuales en Midjourney. Emulando a Louis I. Kahn, *EGA Expresión Gráfica Arquitectónica,* 28(49), 238–251*. https://doi.org/10.4995/ega.2023.19294.

Momot, I. (2022). *Artificial Intelligence in Filmmaking Process Future Scenarios*. [Bachelor's thesis]. https://urn.fi/URN:NBN:fi:amk-2022052712497

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8). https://doi.org/10.1073/pnas.2120481119

Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. *ACM International Conference Proceeding Series*, 192–202. https://doi.org/10.1145/3569219.3569352

Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour and Information Technology*. https://doi.org/10.1080/0144929X.2023.2286532

Parr, M. (1997). Martin Parr's official website. https://www.martinparr.com/

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). *A Primer in BERTology: What we know about how BERT works. Arxiv. Cornell University.* https://doi.org/10.48550/arXiv.2002.12327

Schetinger, V., Di Bartolomeo, S., El-Assady, M., McNutt, A., Miller, M., Passos, J. P. A., & Adams, J. L. (2023). Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models. *Computer Graphics Forum*, *42*(3), 423–435. https://doi.org/ 10.1111/CGF.14841

Son, J.-W., Han, M.-H., & Kim, S.-J. (2019). Artificial Intelligence-Based Video Content Generation. *Electronics and Telecommunications Trends.* https://doi.org/10.22648/ETRI.2019.J.340304

Sosa, R., & Gero, J. S. (2016). Multi-dimensional creativity: A computational perspective. *International Journal of Design Creativity and Innovation*, *4*(1), 26–50. https://doi.org/10.1080/21650349.2015.1026941

Steinfeld, K. (2023). Clever little tricks: A socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, *21*(2), 211–241. https://doi.org/10.1177/14780771231168230

Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards Real-World Blind Face Restoration with Generative Facial Prior. *Arxiv. Cornell University.* https://doi.org/10.48550/arXiv.2101.04061

Zhang, C., & Peng, Y. (2018). Stacking VAE and GAN for Context-aware Text-to-Image Generation. *2018 IEEE 4th International Conference on Multimedia Big Data, BigMM 2018*. https://doi.org/10.1109/BIGMM.2018.8499439

Zhang, L., Chen, Q., Hu, B., & Jiang, S. (2020). Text-Guided Neural Image Inpainting. *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, 1302–1310. https://doi.org/10.1145/3394171.3414017