

The influence of audiovisual elements on the realism of generative AI videos: the case of Sora

Alberto Sanchez-Acedo¹, Alejandro Carbonell-Alcocer^{1,*}, Pasquale Cascarano², Shirin Hajahmadi³, Giacomo Vallasciani², Manuel Gertrudix¹ and Gustavo Marfia²

¹Department of Audiovisual Communication and Advertising, Rey Juan Carlos University, Camino del Molino, 5, 28942 Fuenlabrada, Madrid, Spain

²Department of the Arts, University of Bologna, Via Barberia 4, 4013 Bologna, Italy

³Department of Computer Science and Engineering, University of Bologna, Via Mura Anteo Zamboni 7, 40126 Bologna, Italy

Abstract

Generative Artificial Intelligence (Gen-AI) tools are in the spotlight in every professional field. In the last decade, artificial intelligence technologies that are capable of creating content in various formats such as texts, images, audios, or videos, have emerged. Among the well known tools are those developed by OpenAI, such as ChatGPT, DALL-E and Sora. They can generate text, images, and videos respectively, with the help of instructions given in the form of prompts, in an accessible and efficient way. This study aims to evaluate the attraction, composition and realism of Gen-AI videos in comparison to real videos. Therefore, a quasi-experimental design is conducted using a validated survey with two groups. The experimental group contains two videos produced by Sora as stimuli, while the control group contains two real videos. The results highlight key factors influencing perceived realism, such as natural lighting, saturation, color and perspective. However, the videos that Sora can generate have such a great degree of realism in terms of audiovisual composition that it will be necessary to educate people on the subject of content generation with artificial intelligence to prevent disinformation.

Keywords

Artificial Intelligence, Sora, Videos generated with AI, Text-to-video, Audiovisual analysis, Experiment,

1. Introduction

Generative Artificial Intelligence (Gen-AI) is a specialised field of Artificial Intelligence (AI) that deals with the generation of human-like texts, the creation of images from written descriptions and the production of videos based on predefined instructions [1]. Today, the potential of these Gen-AI tools is the subject of considerable debate on various key issues, ranging from the quality and authenticity of the content created to the ethical implications of their use [2, 3, 4, 5]. However, the ability of Gen-AI to produce original materials in various forms has made a significant impact in various sectors such as creative industries, manufacturing, design, entertainment, and education [2, 5, 6].

Many researchers working for companies and academia have focused their efforts on developing efficient and accessible Gen-AI tools for content creation. Most notably, OpenAI's GPT series, which began with its first release in 2018 and was followed by GPT-2, GPT-3 and GPT-4, as well as its conversational variant, ChatGPT, have significantly impacted the landscape of text generation [1, 7]. GPT is built on the principles of Large Language Models (LLMs) [8] which are designed to process and generate natural language text. The outstanding performances of LLMs, in synthesizing complex information, whether in the form of text or images, stems from the use of advanced techniques such as positional encoding and attention mechanisms [8]. Moreover, the main core of LLMs are complex neural network architectures like Transformers which represent the state-of-the-art for numerous natural language tasks [9, 10].

Later, in 2021, OpenAI continued to push the boundaries of generative AI by releasing DALL-E, a tool capable of generating images based on textual descriptions [11]. While GPT focuses on generating coherent and contextually relevant texts based on input prompts, DALL-E integrates linguistic and

International Workshop on Artificial Intelligence and Creativity (CREAI), co-located with ECAI 2024.

*Corresponding author.



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

visual information and extends this capability to visual content generation [11]. The tool employs the same Transformer architecture as GPT-3 [11]. Unlike traditional models that handle either text or images, DALL-E is a multimodal model [12, 11], meaning that it can understand both types of data, integrating them in creative ways.

The achievements reached by GPT and DALL-E have significantly influenced the text-to-video domain, culminating in the remarkable capabilities showcased by OpenAI's Sora [13, 14, 15] realised in 2024. Sora is an AI model capable of creating realistic and imaginative scenes from text instructions [16, 14]. Similarly to GPT and DALL-E, Sora can analyze text and understand intricate user directives. The process of generating videos is based on a diffusion Transformer architecture [9, 10, 17]. This process begins with a video resembling static noise and progressively refines it by removing the noise over many steps introducing details based on the provided text prompt [17]. Sora is notable for its capability to create up to 1-minute long videos, ensuring a strict adherence to user text instructions while delivering high visual quality and maintaining strong visual coherence, thus allowing users to provide visual contents from, even complex, text narratives [13, 16]. The results of the productions made with Sora are highly realistic and applicable to a multitude of professional fields [15, 14, 18] using a prompt that can be as specific and detailed as the user wishes.

These outstanding results might become controversial. On the one hand, these Gen-AI tools offer tremendous benefits, such as enhancing creativity [19]. On the other hand, since a false perception of the world is possible by making indistinguishable what is real from what has been produced with Gen-AI tools [20, 21], they also raise concerns about misinformation leading to the spread of fake news. Therefore, social awareness is necessary for the verification of exposed stimuli [22], as well as personal judgement in recognising AI-generated materials [23]. From an ethical point of view, the use of these models can lead to privacy violations, as they might inadvertently reveal sensitive information embedded in their training data. Furthermore, they can perpetuate and even amplify biases present in their training datasets, leading to unfair or discriminatory outcomes [24, 25, 26]. The deployment of Gen-AI thus demands careful consideration of these factors, emphasizing transparency, accountability, and the implementation of robust ethical guidelines to mitigate potential harms [27].

This study aims to evaluate the attraction, composition, and realism of Gen-AI videos compared to real videos. According to the manual written by Achi [28], multimedia data must adhere to some standards in terms of audiovisual recordings, such as lighting, colour, or scale, which are the most relevant attributes for visual realism [29]. As a case study, we focus on two landscape videos available on Sora's website. The first video is a recreation of Santorini (Greece), while the second one showcases the Amalfi Coast (Italy). Additionally, we consider two real videos reporting the same content. Through a detailed survey, the study measures attraction via parameters such as illumination, saturation, colourfulness, brightness, and sharpness. Composition is assessed by evaluating the video quality, the presence of shadows, focus, perspective, and shot range. Furthermore, the level of realism of the Sora videos is assessed by determining if the videos appear natural, contain fine details, and resemble drone footage, as well as whether respondents recognize the location and believe the video to be real. Finally, we identify which aspects of attraction and which compositional elements most significantly affect the perceived realism of the Gen-AI videos. For these reasons, we seek answers to the following research questions:

- R.Q.1. How do respondents perceive the attraction and composition of AI-generated videos compared to real videos depicting the same landscapes and environments?
- R.Q.2. What are the key attraction and composition elements that influence the perceived realism of Gen-AI videos of landscapes?

The paper is organized as follows. Section 2 outlines the methodological design of the survey conducted. In Section 3, we analyze the obtained results. Finally, Section 4 discusses these results, offering valuable insights into the research questions and concluding with a discussion of the study's limitations.

2. Research methodology

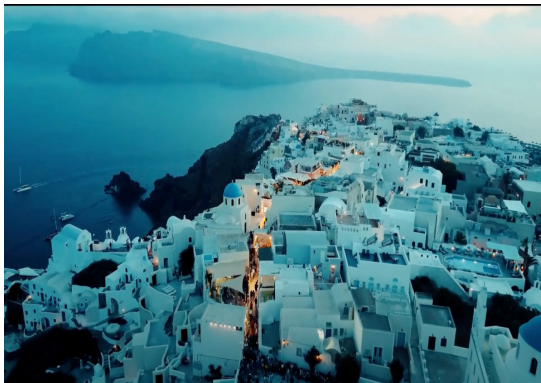
The research objective is to evaluate the attraction, composition and realism of Gen-AI videos compared to real videos, and therefore a quasi-experiment will be conducted. For this reason, an ad hoc survey including either real or Gen-AI videos of landscapes, has been developed. Since the possibilities of producing videos using Gen-AI tools are still limited and the release of Sora is imminent, two landscape videos have been selected from those available on the Sora website. The methodological approach of this research is to make a first approach in the field of Generative AI in video generation and to study how it affects its use in university students. The first is a recreation of Santorini (Greece) and the second shows the Amalfi Coast (Italy). For the selection of the real videos, a search was carried out on Youtube, selecting those that are in the same location and have similar audiovisual characteristics [30, 31]. All videos were customised to have the same duration and resolution. Figure 1 shows a frame for each video. In the validation process, the experts considered the videos to be similar in both similarity and format.



(a) Amalfi Coast (real)



(b) Amalfi Coast (Sora)



(c) Santorini (real)



(d) Santorini (Sora)

Figure 1: Selected frames of real and Sora's videos. (a) Frame of a real video of the Amalfi Coast. (b) Frame of a Sora's video of the Amalfi Coast. (c) Frame of a real video of Santorini. (d) Frame of a Sora's video of Santorini.

The design of a quasi-experiment is based on the construction of two groups, a control group and an experimental group, to which a stimulus is exposed. [32] The quasi-experiment design is based on the collection of information by means of a self-administered online survey [33]. To ensure that the quasi-experiment design fits the research objective, it undergoes validation by expert judges (n=12). Experts in the field of computer science communication and artificial intelligence were selected for this purpose. They were provided with a guide explaining in detail the procedure and method of the experiment, as well as the questions in the questionnaire for collecting information. The purpose of this process is to ensure that there is a degree of agreement in terms of univocity and relevance. [34]. The validation includes the procedure of the quasi-experiment and the questionnaire administered.

The questionnaire (see Table 1) is designed as an information collection system based on the validated framework proposed in [29] for the human characterisation of visual realism in images. Since the questionnaire focuses on variables related to videos, the structure of the questions is modified accordingly, while the variables related to realism, attraction, and composition are maintained. The questionnaire is structured in two sections. The first section is designed to collect socio-demographic variables (P1-P5). The second section contains two videos to be evaluated independently in terms of realism, attraction and composition. In particular, the questions P6-P11 aim at evaluating the level of attraction, P12-P15 address the key elements of composition and, finally, P16-P20 focus on realism. The concepts covered by the survey were explained before the questionnaire was carried out in order to reduce bias in the interpretation of the questions.

Table 1: Survey questions and variables

Question ID	Items	Variable
Section 1: Sociodemographic		
Q1	How old are you?	Age
Q2	Gender. How do you identify?	Gender
Q3	Are you currently working?	Professional activity
Q4	If yes, what is your current position?	Professional activity
Q5	Which country are you from?	Geographical
Section 2: Survey		
Q6	How does the illumination appear to you? (1) Natural (2) Slightly natural (3) Not clearly natural or unnatural (4) Slightly unnatural (5) Unnatural	Attraction
Q7	How does the saturation appear to you? (1) Very saturated (2) Fairly saturated (3) Neutral (4) Slightly saturated (5) Without saturation	Attraction
Q8	How does the colour appear to you? (1) Very colourful (2) Slightly colourful (3) Neutral (4) Slightly uncolorful (5) Uncolorful	Attraction
Q9	How does the brightness appear to you? (1) Very bright (2) Fairly bright (3) Neutral (4) Slightly bright (5) Without bright	Attraction

Question ID	Items	Variable
Q10	How does the sharpness appear to you? (1) Very sharp (2) Moderately sharp (3) Neither sharp nor blurry (4) Moderately blurry (5) Very blurry	Attraction
Q11	What's the quality of the video? (1) High quality (2) Moderately high quality (3) Medium quality (4) Moderately low quality (5) Very low quality	Attraction
Q12	Do you see shadows in the image? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Attraction
Q13	Does the video appear to have objects well focused? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably not (5) Definitely not	Composition
Q14	Does the perspective of the video appear natural? (1) Definitely natural (2) Moderately natural (3) Not clearly natural or unnatural (4) Moderately unnatural (5) Definitely unnatural	Composition
Q15	Does the video appear to be a close-range shot or distant view shot? (1) Very close range (2) Moderately close range (3) Between close and distant (4) Moderately distant view (5) Very distant view	Composition
Q16	Do you recognize the location of the video? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Realism
Q17	Does the colour in the video appear natural? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Realism

Question ID	Items	Variable
Q18	Does the image contain fine details? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Realism
Q19	Does this video look like it is a video taken by a drone? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Realism
Q20	Do you think the video is real? (1) Definitely yes (2) Probably yes (3) Not clearly yes or no (4) Probably no (5) Definitely not	Realism

As the chosen method is a quasi-experimental design, the survey is carried out on two groups. The survey for the control group contains two videos of real landscapes recorded by a professional as stimuli, while the survey for the experimental group contains two landscape videos produced by Gen-AI as stimuli. A non-probabilistic sample is selected as the results of the study aim to collect data to get more insights about the phenomenon of video production with Gen-AI tools. University students with a background in visual arts, theatre and music are therefore taking part in the study. Allocation to the individual groups was randomised and in proportion to each other. The data was collected in April 2024. Data was collected from n=62 participants, 28 from the control group and 34 from the experimental group. An online survey was used for data collection. All participants are young university students in the field of computer science, communication and new technologies.

3. Results

In this section, we report some statistics derived from the survey for both the control and experimental groups in order to seek for answers to the research questions R.Q.1 and R.Q.2 outlined in Section 1. The answers to the socio-demographic questions (Q1, Q2, Q3, Q4, Q5 in Table 1) show that the average age of the participants is 21 years, of which 68% are female and 27% are male. The 81% of respondents are not employed and 71% of them are Italian. Concerning the variables attraction, composition, and realism, the results are presented below in percentages, distinguishing between real videos and videos generated with Sora.

We first focus on the attraction variable by analyzing the answers to Q6-Q12. We found significant differences in terms of distribution between real and AI videos when evaluating the "illumination" (Q6). The bar plots are shown in Figure 2. Participants consider this item, in the case of the Sora's Santorini video, to be predominantly unnatural or slightly unnatural (91%) in the case of the Santorini video by Sora, while in the case of real Santorini video, the lighting is considered more natural and only 32% of respondents consider it slightly unnatural.

For the rest of the items, such as saturation (Q7), colour (Q8), brightness (Q9) or sharpness (Q10), the differences between the real videos and Sora are more neutral and less significant. In the case of saturation, for example, participants mostly considered the four videos to be quite saturated (50% for

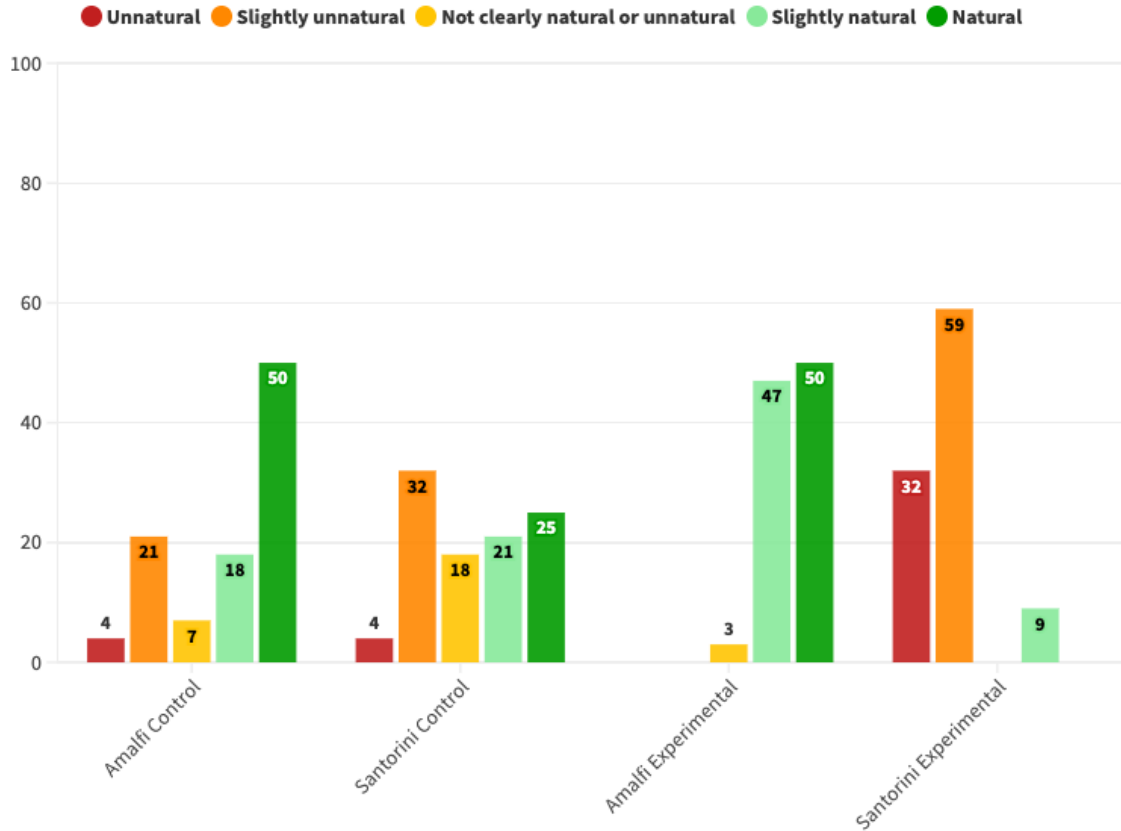


Figure 2: The bar plots depict the responses gathered for question Q6 which assesses the factor of “illumination” concerning the variable of attraction.

the real video of Santorini and 44% for the Sora’s one). The same applies to colour, where participants mostly rated the videos as slightly colourful, regardless of whether it was a real or AI-generated video (57% in the case of the real Amalfi video; 50% in the case of the real Santorini video; 71% in the case of the AI Amalfi video and 47% in the case of the AI Santorini video). In terms of sharpness, the majority of participants felt that all videos were neither sharp nor blurred. Regarding the quality of the videos (Q11), most of them are categorised as of medium quality, with no significant differences between the distribution of the Sora’s videos and the real ones. In summary, if we compare the Sora and real videos, in the case of Santorini there are major differences, especially in the lighting, and less in the saturation or colour elements. We point out that Sora’s video of Santorini stands out as very colourful and saturated compared to the other stimuli. In the case of Amalfi, the differences of the distribution between Sora and real videos in terms of saturation, colour and lighting are small.

We now focus on the composition variable by analyzing the answers to Q13-Q16 which assess the degree of focusing, the perspective and the camera distance. Concerning audiovisual features such as the degree of focusing (Q13), the majority of participants consider real videos better focus if compared to Sora’s ones. More precisely, the percentage of participants perceiving the videos well focused are: 57% for the real video of Amalfi; 64% for the real video of Santorini; 41% in the video of Sora of Amalfi and 47% in the video of Sora of Santorini. Another important compositional element that was analyzed is perspective (Q14). The barplots are shown in Figure 3. For the real videos, most of the respondents believe that the perspective appears natural. For the Sora videos, however, the range of answers is wider, which indicates that the perspective can be perceived as neither natural nor unnatural. Finally, concerning the camera distance (Q16) all the videos are perceived with a moderately or very distant view.

By analyzing the answers to the questions about the realism variable (Q16, Q17, Q18, Q19, Q20),

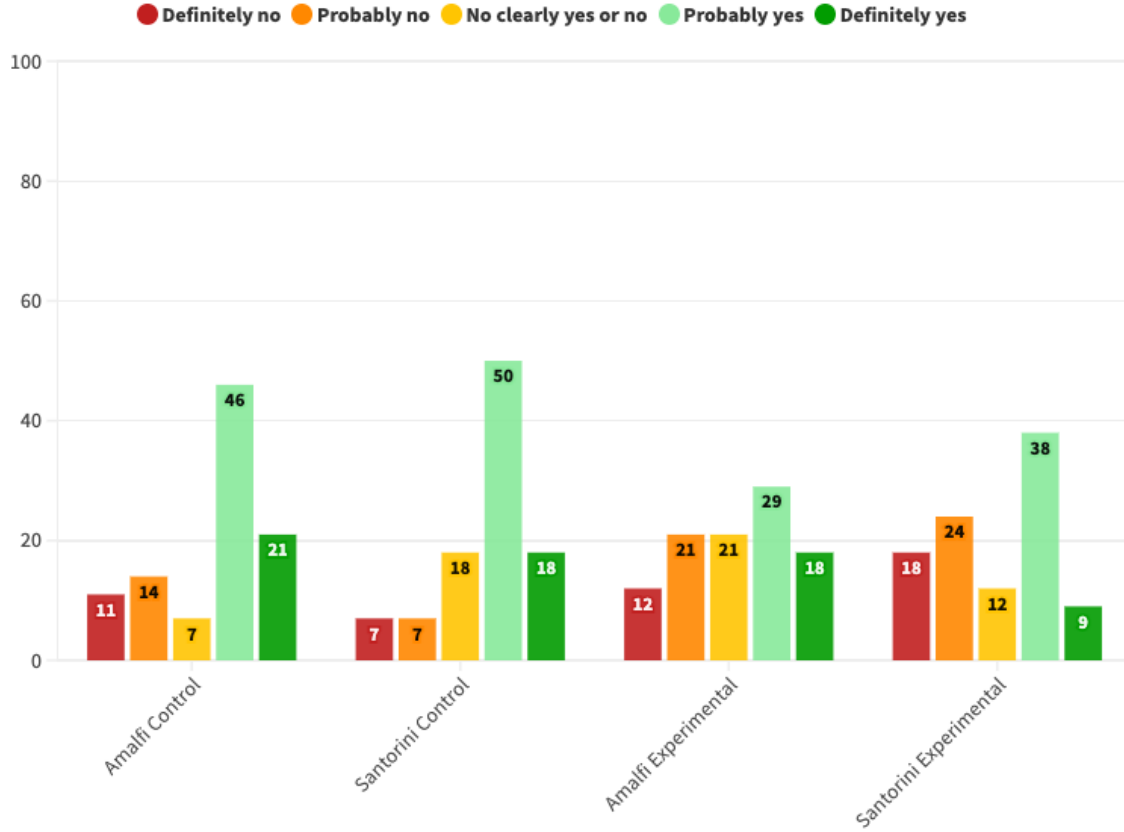


Figure 3: The bar plots depict the responses gathered for question Q14 which assesses the factor of "perspective" concerning the variable of composition.

it turns out that the majority of participants recognise Sora's video about Santorini as false (35%) or probably false (32%). In the case of the AI generated video of Amalfi, participants predominantly (32%) recognise it as probably true. As with the real videos shown in the control group, the majority of participants recognise the Santorini video as probably true (43%). For the Amalfi video, 36% of participants think it is probably false and 32% think it is probably true (Q20). The results are shown in Figure 4. For both the real video of Amalfi and its AI-generated counterpart, participants mostly do not recognize the location. In contrast, participants do recognise the location of Santorini for the most part in both experimental groups (Q16). The results are shown in Figure 5.

4. Discussion and Conclusions

The results of the experiment are based on the assumption that Gen-AI tools for video generation, in particular Sora, are capable of generating realistic content that is practically impossible to differentiate from a real video [35].

In order to answer to R.Q.1. *"How do respondents perceive the attraction and composition of AI-generated videos compared to real videos depicting the same landscapes and environments?"*, in the Section 3 we reported the results obtained by analyzing some items of the attraction and composition variables. The results reveal distinct perceptions of attraction and composition between AI-generated and real videos, thus highlighting both technological limitations and areas of potential improvement in AI video generation. Concerning the attraction variable, we observed the lighting in AI-generated videos, particularly Sora's Santorini video, was largely deemed unnatural, in contrast with the more natural lighting perceived in real videos. Despite this, other attributes such as saturation, colour, brightness, and sharpness showed less pronounced differences, indicating that AI-generated videos can achieve a

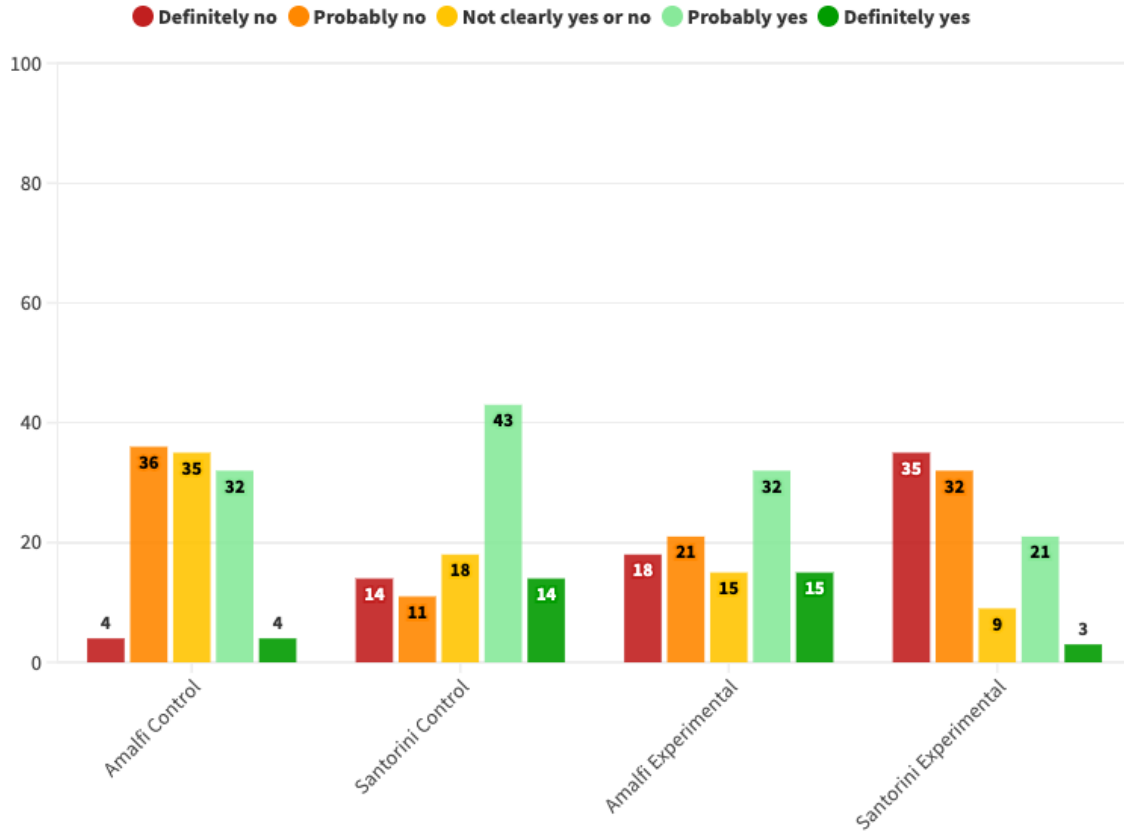


Figure 4: The bar plots depict the responses gathered for question Q20 which assesses the realism.

comparable aesthetic quality in these areas. These results suggest that Sora can replicate certain visual aspects, but it can struggles with replicating natural lighting, which is a critical component of realism [29]. Concerning the composition, the main differences between AI-generated videos and real videos were observed regarding the perspective item. The perspective in AI videos was also perceived less consistently, often seen as neither entirely natural nor unnatural.

We now try to seek answers to R.Q.2. *"What are the key attraction and composition elements that influence the perceived realism of Gen-AI videos of landscapes?"*. It is evident that attraction variable influences the perceived realism of Sora videos. The results indicate that participants could easily distinguish the AI-generated Santorini video due to its unnatural illumination, despite slight differences in saturation and colour. This highlights that elements such as illumination, saturation, and colour are key factors in recognizing an AI-generated video. Conversely, when these elements are closely matched between real and AI-generated videos, as seen with the Amalfi videos, it becomes more challenging for participants to identify the AI-generated content. In these cases, the majority of participants did not recognize the Sora-generated videos as artificial, suggesting that a high degree of similarity in these compositional and attraction elements can enhance the perceived realism of AI-generated videos. The key compositional elements influencing the perceived realism of Gen-AI videos of landscapes include the perspective. Specifically, the natural appearance of perspective in real videos contrasts with the broader range of perceptions for AI-generated videos, where perspectives were often seen as neither natural nor unnatural. Overall, while AI-generated videos are making strides in matching the visual appeal of real videos, significant challenges remain in achieving complete realism, particularly in aspects like lighting and focus that contribute heavily to the perceived naturalness of a scene. These insights underscore the importance of further advancements in AI video synthesis to enhance the authenticity and visual coherence of generated content.

In terms of location recognition, Sora is able to generate highly realistic videos that resemble real

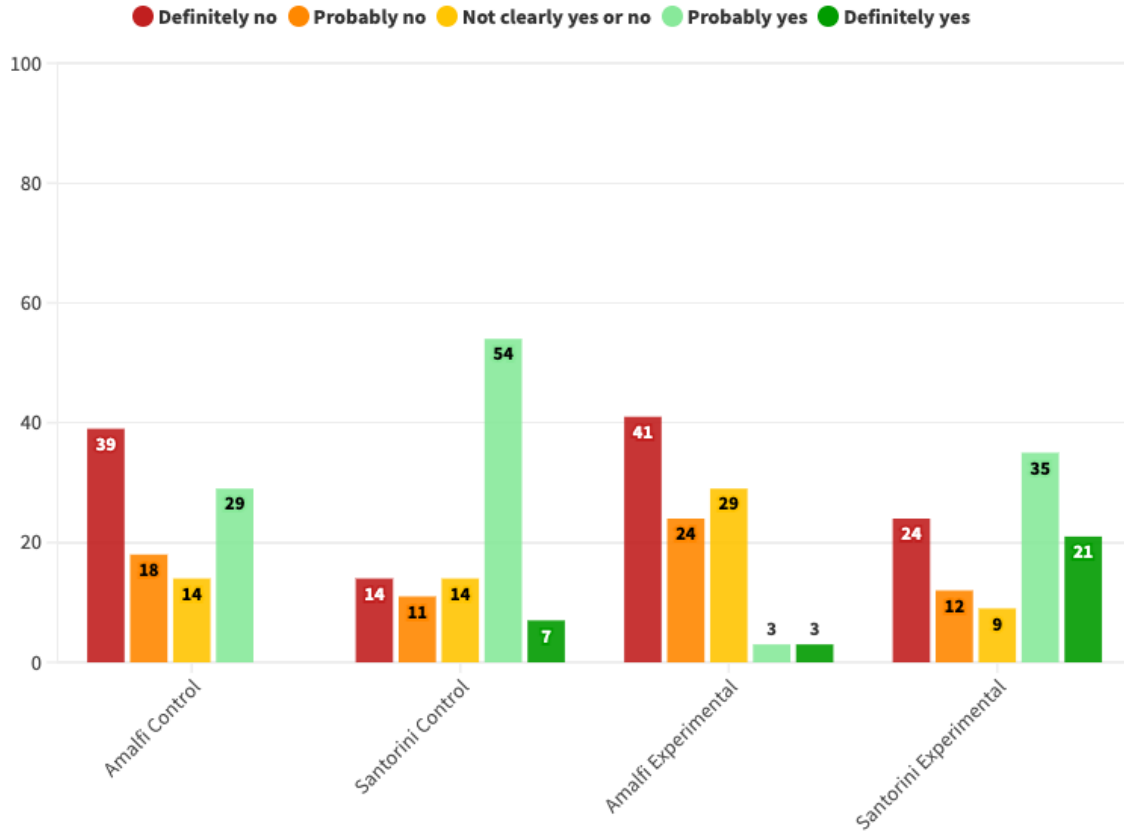


Figure 5: The bar plots depict the responses gathered for question Q16 which assesses the realism.

locations, as demonstrated in the cases of Santorini and Amalfi [21, 36]. Participants were more likely to recognize Santorini because iconic elements, such as the blue domes, were accurately recreated. In contrast, Amalfi lacks such iconic features, making it less recognizable. The era of Sora has just begun and this initial research on the realism of its videos clearly indicates that they are already difficult to distinguish from reality. However, this research suggests that for video generation to achieve results that closely mimic reality, factors such as attraction and composition must be considered to increase the level of realism. Additionally, there is a need to educate viewers on the potential for AI-generated videos. Consequently, regulations have been developed to control such content, and various studies on Sora emphasize the importance of addressing ethical risks related to misinformation[18, 37].

This study faces several limitations. First of all, the accessibility of the Sora tool. At the time of the experiment, Sora was not available to the general public and the study had to rely on default videos provided by the tool, without the ability to explore its full capabilities through detailed instructions. Furthermore, the study exclusively utilized the Sora tool, as other text-to-video AI tools were deemed less effective in producing realistic results. Consequently, Sora is currently regarded as the most efficient tool for generating highly realistic videos using artificial intelligence. Since there is no research of this kind, this experiment is a first approximation to the work of creating videos with AI tools and the results are not generalisable to the rest of the population. To strengthen the findings of this research, replication of the experiment with a larger and more diverse sample of participants across various educational and professional backgrounds is necessary, as well as to compare and study other types of videos that are generated with the Sora tool.

Funding

This work was supported by the Autonomous Community of Madrid (Spain) with a grant for industrial doctorates (IND2022/SOC-23503) with the collaboration agreement with Prodigioso Volcán S.L; Universidad Rey Juan Carlos (ID 501100007511) with a grant call for Personnel in Training 2020 (PREDOC 20-008). This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (Piano Nazionale di ripresa e resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.3 - D.D. 1551.11-10-2022, PE000000004).

References

- [1] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, L. Sun, A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt, arXiv preprint arXiv:2303.04226 (2023).
- [2] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, L. Chen, Generative ai and chatgpt: Applications, challenges, and ai-human collaboration, 2023.
- [3] F. E. Babl, M. P. Babl, Generative artificial intelligence: Can chatgpt write a quality abstract?, *Emergency Medicine Australasia* 35 (2023) 809–811.
- [4] J. Joosten, V. Bilgram, A. Hahn, D. Totzek, Comparing the ideation quality of humans with generative artificial intelligence, *IEEE Engineering Management Review* (2024).
- [5] Z. Epstein, A. Hertzmann, I. of Human Creativity, M. Akten, H. Farid, J. Fjeld, M. R. Frank, M. Groh, L. Herman, N. Leach, et al., Art and the science of generative ai, *Science* 380 (2023) 1110–1111.
- [6] E. A. Alasadi, C. R. Baiz, Generative ai in education and research: Opportunities, concerns, and solutions, *Journal of Chemical Education* 100 (2023) 2965–2971.
- [7] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., Gpt-4 technical report, arXiv preprint arXiv:2303.08774 (2023).
- [8] K. S. Kalyan, A survey of gpt-3 family large language models including chatgpt and gpt-4, *Natural Language Processing Journal* (2023) 100048.
- [9] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, arXiv preprint arXiv:2303.18223 (2023).
- [10] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 2020, pp. 38–45.
- [11] J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, et al., Improving image generation with better captions, *Computer Science*. <https://cdn.openai.com/papers/dall-e-3.pdf> 2 (2023) 8.
- [12] M. Suzuki, Y. Matsuo, A survey of multimodal deep generative models, *Advanced Robotics* 36 (2022) 261–278.
- [13] OpenAI, Creating video from text sora is an ai model that can create realistic and imaginative scenes from text instructions. (2024). URL: <https://openai.com/index/sora/>.
- [14] Y. Liu, K. Zhang, Y. Li, Z. Yan, C. Gao, R. Chen, Z. Yuan, Y. Huang, H. Sun, J. Gao, et al., Sora: A review on background, technology, limitations, and opportunities of large vision models, arXiv preprint arXiv:2402.17177 (2024). URL: <https://doi.org/10.48550/arXiv.2402.17177>.
- [15] R. Sun, Y. Zhang, T. Shah, J. Sun, S. Zhang, W. Li, H. Duan, B. Wei, R. Ranjan, From sora what we can see: A survey of text-to-video generation, arXiv preprint arXiv:2405.10674 (2024).
- [16] T. Brooks, B. Peebles, C. Holmes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, C. Ng, R. Wang, A. Ramesh, Video generation models as world simulators. openai. (2024). URL: <https://openai.com/research/video-generation-models-as-world-simulators>.
- [17] W. Peebles, S. Xie, Scalable diffusion models with transformers, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4195–4205.

- [18] A. J. Adetayo, A. I. Enamudu, F. M. Lawal, A. O. Odunewu, From text to video with ai: the rise and potential of sora in education and libraries, *Library Hi Tech News* (2024). URL: <https://doi.org/10.1108/LHTN-02-2024-0028>.
- [19] A. R. Doshi, O. Hauser, Generative artificial intelligence enhances creativity, Available at SSRN (2023).
- [20] J. Fernández Mateo, et al., Realidad artificial. un análisis de las potenciales amenazas de la inteligencia artificial. (2023).
- [21] R. H. Mogavi, D. Wang, J. Tu, H. Hadan, S. A. Sgandurra, P. Hui, L. E. Nacke, Sora openai's prelude: Social media perspectives on sora openai and the future of ai video generation, *arXiv preprint arXiv:2403.14665* (2024). URL: <https://doi.org/10.48550/arXiv.2403.14665>.
- [22] J. E. Suárez-Roca, G. L. Vélez-Bermello, Verificación de los hechos: Aplicación metodológica en el medio de comunicación el bacán, *Revista Científica Arbitrada de Investigación en Comunicación, Marketing y Empresa REICOMUNICAR*. ISSN 2737-6354. 5 (2022) 163–184. URL: <https://doi.org/10.46296/rc.v5i9.0042>.
- [23] C. Belloch, Las tecnologías de la información y comunicación en el aprendizaje, *Departamento de Métodos de Investigación y Diagnóstico en Educación*. Universidad de Valencia 4 (2012) 1–11. URL: <https://bit.ly/468T21C>.
- [24] A. Ara, A. Ara, Exploring the Ethical Implications of Generative AI, IGI Global, 2024.
- [25] B. Obrenovic, X. Gu, G. Wang, D. Godinic, I. Jakhongirov, Generative ai and human–robot interaction: implications and future agenda for business, society and ethics, *AI & SOCIETY* (2024) 1–14.
- [26] K. Wach, C. D. Duong, J. Ejdy, R. Kazlauskaitė, P. Korzynski, G. Mazurek, J. Paliszkiwicz, E. Ziemia, The dark side of generative artificial intelligence: A critical analysis of controversies and risks of chatgpt, *Entrepreneurial Business and Economics Review* 11 (2023) 7–30.
- [27] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation, *Information Fusion* 99 (2023) 101896.
- [28] M. C. R. Achi, *Manual de Formación Audiovisual*, Cholsamaj Fundacion, 2004.
- [29] S. Fan, T.-T. Ng, B. L. Koenig, J. S. Herberg, M. Jiang, Z. Shen, Q. Zhao, Image visual realism: From human perception to machine computation, *IEEE transactions on pattern analysis and machine intelligence* 40 (2017) 2180–2193. URL: <https://doi.org/10.1109/TPAMI.2017.2747150>.
- [30] R. Shirley, Top 10 places on the amalfi coast - 4k travel guide (2021). URL: <https://www.youtube.com/watch?v=Mupom-sgjAU>.
- [31] G. P. Pro, Santorini, greece - 4k uhd drone video (2021). URL: <https://www.youtube.com/watch?v=rXlqSYZOGnQ&t=456s>.
- [32] C. A. R. Galarza, Diseños de investigación experimental, *CienciAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica* 10 (2021) 1–7. URL: <http://dx.doi.org/10.33210/ca.v10i1.356>.
- [33] C. Ramos-Galarza, Editorial: Diseños de investigación experimental. *cienciamérica*, 10 (1), 1-7, 2021.
- [34] J. Escobar-Pérez, Á. Cuervo-Martínez, Validez de contenido y juicio de expertos: una aproximación a su utilización, *Avances en medición* 6 (2008) 27–36. URL: <https://bit.ly/3IlxiDV>.
- [35] T. Brooks, B. Peebles, C. Homes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, et al., Video generation models as world simulators, 2024.
- [36] M. Kustudic, G. F. N. Mvondo, A hero or a killer? overview of opportunities, challenges, and implications of text-to-video model sora, *Authorea Preprints* 10 (2024). URL: <https://doi.org/10.36227/techrxiv.171207528.88283144/v1>.
- [37] J. Cho, F. D. Puspitasari, S. Zheng, J. Zheng, L.-H. Lee, T.-H. Kim, C. S. Hong, C. Zhang, Sora as an agi world model? a complete survey on text-to-video generation, *arXiv preprint arXiv:2403.05131* (2024). URL: <https://doi.org/10.48550/arXiv.2403.05131>.