Artificial Intelligence for Advertising Creativity: Advances and Implications

Nan Zhang

Hairong Li

Department of Advertising + Public Relations

Michigan State University

Abstract

The integration of AI technologies in advertising has fundamentally altered the processes of producing short video ads and image ads. This paper examines how AI is and will be replacing or enhancing advertising creative professionals. Our analysis is focused on the capabilities of AI in creative strategy, creative execution, and creative effectiveness. Additionally, we explore the limitations and potential risks of AI in advertising creativity.

The increasing use of artificial intelligence (AI) technologies in advertising has resulted in some interesting studies of what role AI can play in advertising creativity. For example, Vakratsas and Wang (2020) developed a creative advertising system to generate and test advertising creative ideas based on AI principles. van Noort et al. (2020) conceptualized an automated brand-generated content model that can balance consumer data and brand data as inputs to algorithms to serve both short- and long-term impacts. Deng et al. (2019) developed an AI system to generate personalized ads and compared such ads with generic ads to show more positive responses to personalized ads. Campbell et al. (2021) proposed a framework to understand consumer response to synthetic ads, which are unreal but convincing ads that AI generative models create. These example studies have advanced our knowledge about the use of AI in advertising development, but little is known about the relationship between AI and human intelligence in advertising creative processes. That is, to what degree AI is and will be replacing or enhancing advertising creative professionals? This is an important issue, as it should affect not only the future of the advertising profession but also advertising education, including what kind of knowledge and skills advertising students should possess for a career in the age of AI.

To address this issue, this study examines the role of AI technologies from the perspective of advertising creativity, focusing on some popular AI models in use for advertising development and the companies that offer such AI models. Advertising creativity has been studied in two areas: creative development and creative effectiveness (West et al., 2019). And creative development can be further divided into two areas of creative effort—creative strategy and creative execution. Creative strategy or message strategy generally refers to “what to say” in an advertising campaign, whereas creative execution or creative tactic refers to “how to say it.” (Taylor, 1999). Wei and Jiang (2005) differentiate creative strategy from creative execution, arguing that creative strategy embodies the themes, positioning, and focal point, whereas the execution aims to execute the creative strategy by selecting advertising appeals, copy, and illustration. Based on these creativity concepts, we explore how AI technologies are used in creative strategy, creative execution, and creative effectiveness.

AI and Creative Strategy

The creative strategy for an advertising campaign is often formed out of a thorough understanding of the brand, the target audience, and the market, including competition (Ashley & Tuten, 2015). With the rise of big data, especially user-generated content in social media, AI and machine learning are increasingly used to extract insights from large-scale unstructured, tracking, and network data for descriptive, causal, and prescriptive analyses (Ma & Sun, 2020). Parker et al. (2021) experimented with 60 working creative professionals who developed creative advertising ideas under three conditions: a strong insight, a weak one, and a no-primed-insight control condition. The findings were that insight gave creative professionals a starting point, a motivational angle from which to generate divergent ideas, and from which they could leap to creative ideas. These findings align with the point view of some advertising professionals, as reported in the Advertising Research Foundation (ARF 2020)'s White Paper, A.I. Driven Creative. It stated that advanced algorithms could help identify themes, patterns, and differences in data that can assist humans in developing ad creatives. Still, they cannot automatically come up with great creative ideas. These studies indicate that consumer insights from big data and machine learning are helpful but still not sufficient and that human intelligence remains to play a primary role in the development of creative strategy.

AI and Creative Execution

Creative execution involves the production of specific advertisements that best manifest the creative strategy of an advertising campaign, which includes what creative elements (e.g., subjects, objects, and scenes) are used and what are the structure and sequences in which these creative elements are presented. The creative elements may come from original images, videos, and audio or stock visuals and audio. Our review of the literature, largely from the field of computer science and industry practices, shows that many AI models are available or already used in advertising creative execution and that they are used in different ways.

Creative Modification

Creative modification refers to how AI models are used to alter the appearance of the creative elements in an advertisement, such as changing the color of an image, reframing a video clip, and adding special effects. Previously, such modification tasks would take creative professionals' hours of time using Adobe Photoshop, Illustrator, Premiere Pro, or other similar tools. Now, AI models can complete these tasks at a large scale with increased precision in just seconds, thus allowing creative professionals to spend their time on the tasks that are still impossible for AI, such as monitoring the quality of the AI-produced advertisements and making final selection decisions. AI modification can be done in several aspects of images, videos, and audio, such as color (Dabas et al., 2020), style (Mao et al., 2017), quality (Kasten et al., 2021), and face swapping (Mohammadi & Kalhor, 2021; Nguyen et al., 2019). Instead of creating their own algorithms, advertising creative professionals can take advantage of available AI models that are described below and listed in Table 1.

Image colorization is a process of adding colors to grayscale images or videos using algorithms that learn from data. Image colorization can be used for various purposes, such as restoring historical images, enhancing artistic expression, and improving visual appeal (Dabas et al., 2020). Image colorization can also be applied to advertising development, as it can help create different versions of advertisements with different colors to suit different contexts, audiences, and preferences. However, image colorization is not a simple task, as it requires understanding the semantics and aesthetics of the images or videos and the intended message and emotion of the advertisements. Therefore, image colorization may not always produce satisfactory results, especially when there is ambiguity or inconsistency in the data or the algorithms. In such cases, human intervention or guidance may be needed to ensure the quality and accuracy of the colorization.

There are two types of AI-enabled image colorization: guided and unguided image colorization (Ardizzone et al., 2019). As the name implies, the unguided is an automatic algorithm coloring. In contrast, the guided involves human intervention (or other references) in the coloring process, such as giving a style reference image or assigning a specific area to a specific color. Human and AI agents can have different roles and relationships in the process of image colorization for advertising development. Depending on the type and level of involvement of human and AI agents, we can identify four scenarios of image colorization: human-led, human-guided, AI-led, and AI-guided.

In the human-led scenario, human agents have full control over the image colorization process and use AI models to assist them in tasks such as finding relevant data or applying filters. For example, a human designer may use an AI model to quickly find images that match a certain keyword or theme for an advertisement campaign. In the human-guided scenario, human agents have partial control over the image colorization process and use AI models as partners to collaborate with them in some tasks, such as giving feedback or suggestions. For example, a human designer may use an AI model to generate different color schemes for an advertisement and choose the best one based on their preference or expertise.

In the AI-led scenario, AI agents have full control over the image colorization process and use human agents as sources to learn from them in some tasks, such as providing data or examples. For example, an AI model may use a large dataset of colored images from previous advertisement campaigns to learn how to colorize new images for a similar campaign. In the AI-guided scenario, AI agents have partial control over the image colorization process and use human agents as validators to verify with them in some tasks, such as checking errors or inconsistencies. For example, an AI model may use a style reference image provided by a human agent to colorize an image for an advertisement and then ask the human agent to confirm if the result is satisfactory.

These scenarios illustrate how human and AI agents can be replaced or enhanced by each other in image colorization for advertising development. In general, we can say that human agents are more likely to be replaced by AI agents when the image colorization tasks are simple, repetitive, or objective and that human agents are more likely to be enhanced by AI agents when the image colorization tasks are complex, creative, or subjective. Similarly, we can say that AI agents are more likely to be replaced by human agents when the image colorization tasks require high-level semantic or aesthetic understanding and that AI agents are more likely to be enhanced by human agents when the image colorization tasks benefit from low-level data or algorithm optimization.

Video editing/enhancing AI can deliver high-end outcomes for video upscaling, denoising, deinterlacing, and restoration by information from several frames. An automated video editor can do basic masking, VFX, color correction, generation, and compositing (Song et al., 2021). For example, Runway is a research and design company specializing in using AI to build new types of creative tools. Their automated video editor can achieve web-based editing in real time and allow its users to create video ads without learning editing techniques. Another example is Rosebud AI, an app for converting a face to a different face, a technique also known as "deepfake" when being used for misleading (Campbell et al., 2021). It uses a deep learning algorithm to create a fake image or video by changing the face in the media to whomever you choose. This technique can be used in video ads, in that you may shoot the video with anyone and substitute the face to the desired person. This technique also can be used in online shopping settings, where a customer may swap the model's facial features in a clothes/makeup virtual try-on (Mohammadi & Kalhor, 2021).

Numerous image and video ads appearing on social media platforms are created using such AI tools, fulfilling a need that humans can hardly meet. In addition, various filters are made available for users to modify their content while posting. All of this has facilitated the wide use of creative modification not only for professional-generated content, but also for user-generated content as well. In general, we can say that human agents are more likely to be replaced by AI agents when the video editing/enhancing tasks are simple, repetitive, or objective and that human agents are more likely to be enhanced by AI agents when the video editing/enhancing tasks are complex, creative, or subjective. Similarly, we can say that AI agents are more likely to be replaced by human agents when the video editing/enhancing tasks require high-level semantic or aesthetic understanding, and that AI agents are more likely to be enhanced by human agents when the video editing/enhancing tasks benefit from low-level data or algorithm optimization.

Creative Conversion

Creative conversion is the process of using AI tools to change the modality of the creative elements in an advertisement, such as turning a video into a text summary, a voiceover into a transcript, or a sound into a graphic (Anderson et al., 2018). Creative conversion is useful, as different modalities of ads are preferred on different platforms. For example, text ads may work better on search engines, while video ads may be more engaging on social media. Creative conversion can also help advertisers reach a wider audience by making their ads more accessible and appealing to different preferences and needs. Existing AI models that are outlined below and are listed in Table 2.

One of the challenges of creative conversion is understanding the content and context of the original ad and preserving its meaning and message in the new modality. This requires AI to have a high level of content understanding, which is the ability to assign names for objects, recognize and infer the three-dimensional structure of things, and understand relational emotions, actions, and intentions (Deng et al., 2009). Content understanding has been well developed in text and images, with accuracy rates of about 95% in the ImageNet (the largest computer vision dataset) competitions (Tan et al., 2019). However, videos are much more challenging for AI to understand, even though they are just a series of images. Videos contain temporal sequences, for example, and moving objects may have different semantics. It is difficult to know what kind of dance it is by looking at a group of dance action pictures individually, and it is not easy for AI to recognize the movement and the change of an object. Therefore, video is multi-modal, including images, facial expressions, audio, text, etc. AI needs to comprehensively analyze multiple dimensions, such as audio and images simultaneously to form a more "stereo" cognition (Huang et al., 2021).

To enrich AI's cognition, programmers must create a cognitive system for it, or a knowledge graph. It can be understood as the “memory” of AI, and each concept in the graph is not a simple image but a multi-dimensional data (Ji et al., 2021). A knowledge graph can be decided and built depending on the company and purpose. For example, a video advertisement may have a sentiment from facial expressions, images, music, and audio. In the facial expression dimension, AI can use facial recognition technology to identify emotions such as happiness, sadness, anger, etc. In the image dimension, AI can use object detection and scene recognition technology to identify objects such as cars, animals, buildings, etc., and scenes such as beach, forest, city, etc. In the music dimension, AI can use music analysis technology to identify genres such as rock, pop, classical, etc., and moods such as energetic, calm, romantic, etc. (Fan et al., 2020). A real-life example of creative conversion was a series of Coca-Cola ads where the actors were thirsty and drinking Coke, leaving the audience with a refreshing feeling. To understand the video, AI needs to recognize the actors, their gestures, and facial expressions, and then locate the objects, such as the Coke in an actor’s hand and when they drank it. At the same time, AI also tries to understand the audio track and combine it with the video content to and perceive the video's sentiment better. Such analysis clarifies how AI processes video ads and transforms the modality from video to text.

Creative conversion is not only a technical challenge but also a creative one. AI needs to be able to generate new content that is relevant, coherent, engaging, and persuasive in the new modality. For example, if AI converts a video ad into a text summary, it needs to capture the ad's main message and emotional appeal in a concise and catchy way. If AI converts a voiceover into a transcript, it needs to preserve the tone and style of the speaker and use proper punctuation and formatting. If AI converts a sound into a graphic, it needs to use colors, shapes, and symbols that match the mood and meaning of the sound.

Creative conversion is an emerging field with many potential applications and benefits for advertisers and consumers. It can help advertisers create more diverse, personalized ads that suit different platforms and audiences. It can also help consumers access more information and entertainment in their preferred modalities. Creative conversion is a promising AI research and development direction that can enhance human creativity and communication.

AI-powered creative conversion is not meant to replace human advertising professionals but rather to enhance their capabilities and efficiency. AI can help advertising professionals with tasks that are tedious, time-consuming, or require large-scale data analysis, such as generating multiple variations of ads, testing and optimizing them, and targeting them to the right audiences. AI can also provide insights and suggestions that can inspire human creativity and innovation. However, AI cannot replace the human touch that is essential for creating emotional connections and compelling stories with consumers. Advertising professionals still need to use their intuition, judgment, and experience to craft their campaigns' overall strategy, vision, and message. They also need to monitor and evaluate the performance and impact of their AI-powered ads and adjust as needed. AI is a powerful tool that can augment human intelligence and creativity, but it is not a substitute for it.

Creative Generation

Creative generation refers to the way that AI tools are used to produce advertisements from scratch, using either original or stock visuals and audio, or completely based on algorisms (Brynjolfsson & Mcafee, 2017). AI tools have been developed to write poems (Ruipérez et al., 2017), composing music (Miller, 2019), and painting (Colton, 2012). Advertising creative professionals can leverage the AI models that are currently available, as described below and listed in Table 3. Most popular among the AI generative tools are GANs, which simulate humans' understanding of traditional generative models and achieve satisfactory results in image generation. Such tools have been used to produce synthetic advertisements (Kietzmann et al., 2020), and researcher have proposed frameworks to understand how consumers respond to such "manipulated" advertisements (Campbell et al., 2021).

With the help of advanced generative AI tools, individuals can provide textual descriptions of their requirements and request various generated materials. Recently released tools such as DALL-E2, Midjourney, and Stable Diffusion can create high-quality images based on textual prompts. ChatGPT can interact with users and generate detailed scripts, advertising strategies, and even precise prompts to feed into image-generative applications. These popular AI generative applications use machine learning models to accomplish different tasks.

Generative models, such as GPT-3 and GPT-4, are natural language processing models that use Transformers to understand text input and generate more reliable output based on context. GPT-3, in particular, has been trained on a massive amount of data and can perform various natural language tasks, including text completion, question-answering, and machine translation (Floridi & Chiriatti, 2020). To illustrate how GANs work, consider the process of submitting a paper to an editor. The editor provides feedback, the writer revises the paper, and the cycle continues until the editor finds nothing else to improve. In GANs, the generator is the writer, and the discriminator is the editor. For image generation, the generator creates an image, which the discriminator evaluates. Feedback is then sent to the generator to improve the image until the discriminator cannot differentiate between the generated and real images. GANs can also be used for audio and video generation, although they are not as well developed as image generation.

Generative models for images have a different focus than the GPT family, which primarily generates sequential data such as text. The goal of image generative models is to create a multi-dimensional vector that represents an image. There are four main types of image generative models: Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), Variational Autoencoders (VAEs) (Doersch, 2016), Flow-based generative models (Nalisnick et al., 2018), and Diffusion Models (DMs) (Kingma et al., 2021). GANs work by using a generator, which creates an image, and a discriminator, which evaluates the image and provides feedback to the generator to improve it until it cannot be differentiated from a real image. While GANs are most used for image generation, they can also be used for audio and video generation. VAEs contain an encoder and decoder, with the encoder extracting high-level features of an image and the decoder using those features to generate a new image. Flow-based generative models replace the encoder and decoder of VAEs with transformations and inverse transformations, making them easier to train. Diffusion Models work by adding noise to training data and then reversing the process to recover the original data. Although each generative model has a different architecture and training strategy, their ultimate goal is the same: to generate images with different types of input.

These generative models also have potential applications in advertising. For example, the ChatGPT, which is embedded with GPT-3 and future GPT-4, can generate advertising content, product descriptions, social media posts, and entire articles with simple prompts provided by the user. By quickly and efficiently generating creative solutions for users' questions, these models can help advertisers produce compelling content that effectively promotes their products or services. Powered by the aforementioned generative models, there are plenty of opportunities to create advertisements for different tasks:

Text-to-text is a natural language processing task in which an AI system generates responses to user queries. One of the most popular models used for this task is ChatGPT, which is powered by GPT-3 and will soon be updated to GPT-4. This model can quickly and effectively create advertising content, product descriptions, social media posts, and even entire articles based on simple prompts provided by users. For instance, a user could ask ChatGPT to generate an advertisement for a crispy, sweet, and organic apple. ChatGPT responded “Fade in: We see a close-up shot of a beautiful, bright red apple with a few droplets of water glistening on its skin. The camera pulls back slowly to reveal a sunny orchard with rows upon rows of apple trees stretching out into the distance. Narrator (voiceover): "Introducing the perfect apple. Crispy, sweet, and organic." Cut to a shot of a smiling farmer in overalls, holding a basket of apples and looking at the camera…”

Text-to-Speech is also called speech synthesis, as this technique artificially generates the human voice through text. The voice is a continuous analog signal, and the synthesis process is simulated by a computer and a digital signal (Van Den Oord et al., 2016). This technology allows a piece of text to become a speech and synthesizes voices with different timbres, speech rates, and intonations as needed, making the machine speak like a human. Many applications and free tools can be found online and used in the creative conversion. Recently, Google Ads launched a new feature that allows users to add a voice-over to their YouTube video ads. They believe that the contribution of auditory elements, such as music, sound effects, and voice-over, to attracting people's attention and inducing them to act is substantial (Port, April 2022). Unlike the generative models for text and image, the Text-to-Speech model is particularly trained by speech data. For example, the WaveNet model from Google research extracts the underlying structure of speech, such as tone sequencing and real speech waveform characteristics during the training; once given a text input, a trained WaveNet model can generate the corresponding speech waveform from scratch. This technology has the potential to greatly reduce human errors and efforts, allowing for more efficient and effective advertising campaigns.

Text-to-speech technology can revolutionize the advertising industry by simplifying the process of creating high-quality voiceovers for ads. With AI-generated speech, advertising professionals can create multiple voiceovers with different tones, speaking speeds, and intonations more efficiently than ever before. For instance, in our "apple advertisement," instead of using a human narrator, we can rely on text-to-speech models to generate audio with various languages, speech rates, and even emotions.

Text-to-image technology has also gained popularity in AI, with tools such as DALL-E2, Stable Diffusion, and MidJourney at the forefront. DALL-E2 (Ramesh et al., 2021), released in July 2022, incorporates both Generative Adversarial Networks (GANs) and Diffusion Models (DMs) to create realistic images from textual descriptions and improve their quality. In contrast, Stable Diffusion is an open-source text-to-image model that solely relies on Diffusion Models and allows billions of people to generate stunning artwork in seconds. MidJourney is another widely used text-to-image tool fine-tuned on Stable Diffusion and is available for free testing on Discord.

Liu et al. (2022) predict that AI could eventually replace human illustrators and artists. However, this viewpoint is not universally shared, as different artists and illustrators can interpret the same text prompt uniquely. While AI can quickly generate large-scale, programmatic designs, it may lack human artists' emotional intelligence and creativity.

Nevertheless, technology has the potential to greatly transform the design industry, especially in areas like fashion, beauty, and advertising. For instance, in our “apple advertisement”, we can forego the need for a camera and a physical scene by using a text-to-image model and fine-tuning the production process. To demonstrate the effectiveness of existing AI text-to-image tools, we input ChatGPT's script into three different AI models. The results give us an example of how to combine ChatGPT and the text-to-image tool together to generate high-quality, low-cost demo of the Apple advertisement.

Figure 1. Examples of to combine ChatGPT and the text-to-image tool together to generate high-quality, low-cost demo of the Apple advertisement.

Via Stable diffusion Via Midjourney Via DALL-E

Stable Diffusion prompt: “We see a close-up shot of a beautiful, bright red apple with a few droplets of water glistening on its skin.”
Midjourney prompt: “The camera pulls back slowly to reveal a sunny orchard, with rows upon rows of apple trees stretching out into the distance.”
DALL-E prompt: “Cut to a shot of a smiling farmer in overalls, holding a basket of apples and looking at the camera.”

Text-To-Video can produce a video based on the user’s input. Different from the aforementioned Text-To-Text, Text-To-Speech, and Text-to-Image tools which embedded the generative models (GPT family, WaveNet, GANs, DMs) to create new data, this task requires a database to search and assemble the clips. For example, Write-A-Video can transform themed text to video clips (Wang et al., 2019). When a creative professional provides a script, the tool will identify keywords and search in the candidate shots database before deploying optimization algorithms to assemble the clips. There are two places where AI is used—generating the clip database and optimizing the assembling. To create the database, the tool first collects videos from the internet with all different keywords, then uses object detection to generate more associated keywords. During the assembling process, the tool uses a visual semantic embedding algorithm. It is essentially a keyword and video-matching process. For example, "giraffe has long neck" and "we went to a zoo today, and I saw a giraffe" have different sentiments while both keywords are "giraffe." The AI optimization algorithm learns from a training set and then calculate the cosine similarity between the video and text.

The expanded role of AI techniques in creative execution, including creative modification, conversion, and generation, indicates that AI can play a primary role in producing advertisements, especially for video ads on social media platforms like TikTok, where a huge number of video ads are running each minute (Herrman, 2019). In other words, AI tools have relieved creative professionals from tedious technical tasks so that they can focus on the tasks that require human intelligence, such as previewing, polishing, scheduling, and releasing AI-produced advertisements.

Although machine learning algorithms and computer vision technologies play a critical role, human agents still play important roles in the process. The training datasets that machine learning algorithms use are put together and labeled by people. This includes tasks such as tagging images and videos with labels for objects, actions, emotions, and intentions. Machine learning algorithms can't learn to recognize and understand video and image content without good, labeled datasets. Additional human agents often fine-tune machine learning models and algorithms to improve accuracy and performance. This includes tasks like adjusting the hyperparameters, choosing and putting different machine learning architectures into place, and making the whole training process as efficient as possible. Finally, human agents are needed to interpret and analyze the output of machine learning algorithms to ensure that the results are accurate and relevant to the intended purpose. This includes tasks such as verifying the accuracy of object and facial recognition, evaluating sentiment analysis, and ensuring that the overall interpretation of the content aligns with the intended message or goal. Overall, human agents play a crucial role in developing and deploying video- and image-to-text technologies, working in conjunction with machine learning algorithms to ensure accurate and effective results.

AI and Creative Effectiveness

Advertising creative effectiveness is examined in terms of how consumers react to either the creative advertisements that are judged as high in originality and appropriateness or the elements of a creative advertisement such as types of appeal or novel means of presenting advertising ideas (West et al., 2019). In the past, such advertising creative effectiveness research was normally conducted either in a lab test, which would assess the response of a sample of consumers to a print ad or TV commercial, or a field test that would typically be a post-campaign assessment of the effectiveness of the ad or ads of the campaign. Unlike the conventional copy testing, AI technologies can assess the effectiveness of a set of AI-produced advertisements in real-time, identify the most effective elements or sequence, optimize accordingly to generate another set of advertisements for new placement and further assessment. Another difference is that the criteria of effectiveness that AI often uses in such assessment are primarily behavioral, such as viewability, click-through rates, and conversions, which are all short-term effects. In comparison, conventional research of advertising creative effectiveness tends to include demographic and psychological measures as well (Costa, 2019).

Using AI technologies to assess advertising creative effectiveness is basically an automatic process during an ongoing advertising campaign, where advertising professionals need only to set the effectiveness criteria, adjust the budget based on the campaign performance, and communicate with the client about the campaign progress (Mogaji et al., 2020).

AI and Advertising Creativity

The implementation of AI in advertising creativity varies by the modality of advertisements to be produced. We explore the roles of AI and humans in the production of short video ads and image ads, as they are widely used in social media platforms and ecommerce websites. The demand for large quantities of ads and fast rates of ad variations are increasingly served with AI-enabled automation. Instead of describing the detailed production process, we focus on several key functions of such production that characterize the roles of AI and advertising professionals.

Short Video Ads

In the production of short video ads, a function often called the **elements center** is developed and maintained as a hub of digital assets. These digital assets are decomposed videos from original footage, historical videos from previous campaigns or other client sources, or copyrighted or copyright-free stock videos. The decomposed videos can be frame-by-frame or scene-by-scene with various tagging and ready to use in producing versions of short video ads. The decomposing and tagging tasks are completed with AI algorithms, whereas producing original video and selecting feeds of other videos still rely on advertising professionals.

Video ad maker is another function, which is designed to produce short video ads in the computing cloud using decomposed video assets based on the tags of the audience, the context, and the media outlet. This process was called as "a computational creative process" by the founder of a major AI creative service. In an interview, another executive mentioned that a 5-minute video clip could be decomposed to produce more than 100 different 15-second video ads. In this process, the role of advertising professionals is largely in managing of the production and making ad clearance decisions—all AI-produced short video ads must be previewed by humans before being placed via APIs (application programming interfaces) in various media outlets.

The next function is a **creative data engine**. It can collect the viewer's response to each short video ad in real-time, analyze such responses by second, scene or even frame, and visualize the results for advertising professionals. For example, a line chart for a short video ad over a period of 30 seconds may show the number of impressions peaked at 2 seconds and soon faded and leveled at 15-20% of the audience, indicating about 80-85% of the audience skipped the video ad after initial viewing. For the remaining audience, other lines may show a few small ups in the number of clicks or conversions over the rest length of the video ad. The visualizer allows advertising professionals to inspect what scenes or even frames of the video ads are associated with the peak and small ups over the time axis for the knowledge about effective creative tactics.

Creative optimizer can be part of the creative data engine or a separate function. It is designed to identify effective frames and scenes of a short video ad and tag them with the viewer response, along with information about the viewer, context, and media outlet. Such information can be used for both creative production and media placement. To comply with the data privacy regulation, less demographic but more interest tagging is used for creative optimization. And these optimized assets are then fed back to the elements center and the video ad maker for the next round of creative production.

Image Ads

Similar functions exist in the production of image ads, such as banner ads and sponsorship ads. Banner ads are typically placed in prominent positions to attract the consumer's attention (Chen & Cheung, 2020) and drive traffics (Menon & Soman, 2002). With the numerous products to promote, the design of banner ads can benefit tremendously from AI-enabled automation, including design time minimization (Jauhari et al., 2018) and efficient use of creative templates (Yang, 2019). Automated banner ad generation tends to work the best for high-frequency and fragmented design needs, as the goal of such automation is to attract users to click and maximize transactions at minimal design costs.

Creative templates still need humans to develop, including the selection of creative components for the templates (Chen et al., 2019; Jauhari et al., 2018). Human designers can gain experiences through repetition and intensive design, then their heuristics and procedures can become automatic creation "practices." However, human designers are not able to explain what makes banner ads creative and effective (Fourquet-Courbet et al., 2007). With the help of deep learning and reinforcement learning methods, human designers now can extract features from data and design experience, and create a design knowledge profile for e-commerce to automatically generate banner ads (Zhang et al., 2017).

A typical banner ad has some structured layers—product, image, text, and background (Zhang et al., 2017). To illustrate the major functions of the banner ad production, we use Luban as an example. Luban was developed by Taobao based on big data and user behavior analysis, and it can produce 50 million banner ads each day (Liu et al., 2019). In Luban, human designers only need to choose a theme and upload the product images, along with the campaign objective, then Luban will generate banner ads in several different styles in seconds. The AI-powered application can work automatically on all the aspects of production, such as digital asset analysis, picture retouching, color blending, and layout (Zhang et al., 2017). Luban would generate different copywriting and slogans based on the campaign objective based on its knowledge profile, its core technology that claims to be China's largest database of marketing photographs by thoroughly analyzing millions of marketing visuals (Team, 2018). Luban has built its knowledge profile on features that were extracted by machine learning algorithms and classified into different categories to meet design needs (Chen et al., 2019).

In sum, as for the role of AI and human designers, the powerful Luban system has been developed and maintained by human programmers and used by advertising professionals. It generates banner ads for each advertising campaign based on the information provided by advertising professionals. Luban can save a great deal of time and labor, but the banner ads it creates are not something highly creative though they are adequate for mass personalization to certain degrees.

Summary and Conclusion

Our analysis shows that the integration of AI technologies in advertising creativity has fundamentally altered the process of producing short video ads and image ads, including the creative preparation, creative production, creative analysis, and creative optimization. On one hand, AI has taken over many tedious tasks such as decomposing and tagging video clips, producing ads based on multiple signals at the scale, analyzing and visualizing the viewer responses, and optimizing creative assets. Many of these tasks would be either time-consuming or almost impossible for humans to complete given the reasonable cost and available time for an advertising campaign, especially on social media or e-commerce websites. On the other hand, advertising professionals are still indispensable to manage many tasks that are not easily carried out by AI technologies. In many ways, more creative tasks still demand human intelligence and different skill sets.

Our analysis has focused largely on the capabilities of AI in creative strategy, creative execution, and creative effectiveness, but the limitations of AI and potential risks in advertising creativity must not be overlooked. The most important question for advertising professionals is whether the current AI is perfect and can replace humans. The answer is certainly "No." AI is not perfect. Or as it for now, AI is not yet perfect. For all AI models, they do not have any kind of memory (Tamkin et al., 2021). In other words, it is unable to recall previous inputs or outputs. For example, in a text generation, a sufficiently extended paragraphs, language model GPT-3 may lose consistency, contradict themselves, and occasionally contain irrational phrases or paragraphs.

Additionally, although AI models can compose unique content, they still create copies based on what already exists on the web. As a result, all the ideas presented are essentially reflections on what someone has already written somewhere. The creation is limited by the input data. Besides, AI is not always able to produce high-quality content. Keep in mind that many networks are trained from internet data with unsupervised learning methods. The advantage of unsupervised learning is that data labeling is a tedious and time-consuming task. However, data sets from the internet could have good or bad quality contents, thus resulting in spotty contents. Therefore, even with automated content generation, whether short video ads or image ads, humans must inspect them to ensure all are appropriate to use for campaigns.

Another potential issue is that AI is an algorithm, not an art. Copywriting or image making are not only graphic design, but also conveyance of emotions and feelings. AI relies on algorithms, data, and powerful computing power. Emotional intelligence, or the capacity of computers to identify and understand human emotions and consciousness, is only beginning (De Bruyn et al., 2020). It takes time for AI to gradually master the design logic in understanding of and responding to human emotions in its interactions with humans, including in the field of advertising. Having imagination and understanding human nature is the core of creating high-quality content. However, we are still far from “Strong AI” (Ng & Leung, 2020) at present .

This study has scratched only the surface of the issue of AI and advertising creativity in terms of the role of AI and advertising professionals. As AI technologies are continuously advancing and more advertisers, ad agencies, and ad tech companies are integrated AI technologies in the process of advertising creativity, new issues are emerging and deserve additional research. For example, how can AI help advertising professionals to formulate the creative strategy for a new advertising campaign? What signals can AI use personalize an image ad in a way that the viewer feels highly relevant and not intrusive at all? How can AI produce a video ad that is emotionally congruent with the viewer in the moment? These are just some research questions for future exploration. Also, this study is conceptual, so empirical investigations are necessary to verify how AI technologies are shaping the advertising creativity process, advance the literature on advertising creativity, and find better solutions to meet the challenges of advertising in the age of AI.

References

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. Proceedings of the IEEE conference on computer vision and pattern recognition,

Ardizzone, L., Lüth, C., Kruse, J., Rother, C., & Köthe, U. (2019). Guided image generation with conditional invertible neural networks. *arXiv preprint arXiv:1907.02392*.

Arık, S. Ö., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., Li, X., Miller, J., Ng, A., & Raiman, J. (2017). Deep voice: Real-time neural text-to-speech. International Conference on Machine Learning,

Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, *32*(1), 15-27.

Brynjolfsson, E., & Mcafee, A. (2017). Artificial intelligence, for real. *Harvard Business Review*, *1*, 1-31.

Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2021). Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising. *Journal of Advertising*, 1-17. https://doi.org/10.1080/00913367.2021.1909515

Chen, G., Xie, P., Dong, J., & Wang, T. (2019). Understanding programmatic creative: The role of AI. *Journal of Advertising*, *48*(4), 347-355.

Chen, Z. T., & Cheung, M. (2020). Consumption as extended carnival on Tmall in contemporary China: A social semiotic multimodal analysis of interactive banner ads. *Social Semiotics*, 1-21.

Colton, S. (2012). The painting fool: Stories from building an automated painter. In *Computers and creativity* (pp. 3-38). Springer.

Costa, M. J. C. T. N. (2019). *Advertising effectiveness: the impact of advertising creativity and emotional media context*

Dabas, C., Jain, S., Bansal, A., & Sharma, V. (2020). Implementation of image colorization with convolutional neural network. *International Journal of System Assurance Engineering and Management*, *11*(3), 625-634.

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K.-U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of interactive marketing*, *51*, 91-105. https://doi.org/10.1016/j.intmar.2020.04.007

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition,

Deng, S., Tan, C.-W., Wang, W., & Pan, Y. (2019). Smart Generation System of Personalized Advertising Copy and Its Application to Advertising Practice and Research. *Journal of Advertising*, *48*(4), 356-365. https://doi.org/10.1080/00913367.2019.1652121

Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

Fan, Y., Li, V., & Lam, J. C. (2020). Facial expression recognition with deeply-supervised attention network. *IEEE transactions on affective computing*.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681-694.

Fourquet-Courbet, M.-P., Courbet, D., & Vanhuele, M. (2007). How web banner designers work: The role of internal dialogues, self-evaluations, and implicit communication theories. *Journal of advertising research*, *47*(2), 183-192.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Hao, S., Guo, B., Wang, H., Liang, Y., Yao, L., Wang, Q., & Yu, Z. (2021). DeepDepict: Enabling Information Rich, Personalized Product Description Generation With the Deep Multiple Pointer Generator Network. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *15*(5), 1-16.

Herrman, J. (2019). How TikTok is rewriting the world. *The New York Times*, *10*.

Hongsuck Seo, P., Nagrani, A., Arnab, A., & Schmid, C. (2022). End-to-end Generative Pretraining for Multimodal Video Captioning. *arXiv e-prints*, arXiv: 2201.08264.

Huang, J.-H., Murn, L., Mrak, M., & Worring, M. (2021). Gpt2mvs: Generative pre-trained transformer-2 for multi-modal video summarization. Proceedings of the 2021 International Conference on Multimedia Retrieval,

Jauhari, A., Anamisa, D. R., Putro, S. S., & Rizqi, F. M. (2018). Automatic Application Of Graphic Design To Banner Model. International Conference on Science and Technology (ICST 2018),

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Kasten, Y., Ofri, D., Wang, O., & Dekel, T. (2021). Layered neural atlases for consistent video editing. *ACM Transactions on Graphics (TOG)*, *40*(6), 1-12.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business horizons*, *63*(2), 135-146.

Kingma, D., Salimans, T., Poole, B., & Ho, J. (2021). Variational diffusion models. *Advances in neural information processing systems*, *34*, 21696-21707.

Li, M., Lin, J., Ding, Y., Liu, Z., Zhu, J.-Y., & Han, S. (2020). Gan compression: Efficient architectures for interactive conditional gans. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Liang, J., & Lin, M. C. (2021). Machine learning for digital try-on: Challenges and progress. *Computational Visual Media*, *7*(2), 159-167.

Liu, K.-l., Li, W., Yang, C.-y., & Yang, G. (2019). Intelligent design of multimedia content in Alibaba. *Frontiers of Information Technology & Electronic Engineering*, *20*(12), 1657-1664.

Liu, V., Qiao, H., & Chilton, L. (2022). Opal: Multimodal Image Generation for News Illustration. Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology,

Lv, F., Lu, F., Wu, J., & Lim, C. (2018). MBLLEN: Low-Light Image/Video Enhancement Using CNNs. BMVC,
Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International journal of research in marketing*. https://doi.org/10.1016/j.ijresmar.2020.04.005
Mao, H., Cheung, M., & She, J. (2017). Deepart: Learning joint representations of visual arts. Proceedings of the 25th ACM international conference on Multimedia,
Menon, S., & Soman, D. (2002). Managing the power of curiosity for effective web advertising strategies. *Journal of Advertising*, *31*(3), 1-14.
Miller, A. I. (2019). *The artist in the machine: The world of AI-powered creativity*. Mit Press.
Mogaji, E., Olaleye, S., & Ukpabi, D. (2020). Using AI to personalise emotionally appealing advertisement. In *Digital and social media marketing* (pp. 137-150). Springer.
Mohammadi, S. O., & Kalhor, A. (2021). Smart Fashion: A Review of AI Applications in Virtual Try-On & Fashion Synthesis. *Journal of Artificial Intelligence*, *3*(4), 284.
Nalisnick, E., Matsukawa, A., Teh, Y. W., Gorur, D., & Lakshminarayanan, B. (2018). Do deep generative models know what they don't know? *arXiv preprint arXiv:1810.09136*.
Ng, G. W., & Leung, W. C. (2020). Strong artificial intelligence and consciousness. *Journal of Artificial Intelligence and Consciousness*, *7*(01), 63-72.
Nguyen, T. T., Nguyen, Q. V. H., Nguyen, C. M., Nguyen, D., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573*.
Parker, J., Koslow, S., Ang, L., & Tevi, A. (2021). How Does Consumer Insight Support the Leap to a Creative Idea? *Journal of advertising research*(March), 30-43. https://doi.org/10.2501/JAR-2020-012
Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. *arXiv pre-print server*. https://doi.org/None
arxiv:2102.12092
Ruipérez, C., Gutiérrez, E., Puente, C., & Olivas, J. (2017). New Challenges of Copyright Authorship in AI. Proceedings on the International Conference on Artificial Intelligence (ICAI),
Song, J., Deng, J., Xin, X., Wu, X., & Liu, H. (2021). Exploration of Online and Offline Media Convergence Based on Artificial Intelligence Technology. 2021 IEEE/ACIS 19th International Conference on Computer and Information Science (ICIS),
Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. F. (2020). Learning to summarize with human feedback. *Advances in neural information processing systems*, *33*, 3008-3021.
Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *arXiv preprint arXiv:2102.02503*.
Tan, F., Wei, Z., Pani, A., & Yan, Z. (2019). User response driven content understanding with causal inference. 2019 IEEE International Conference on Data Mining (ICDM),
Tanberk, S., Dağlı, V., & Gürkan, M. K. (2021). Deep learning for videoconferencing: A brief examination of speech to text and speech synthesis. 2021 6th International Conference on Computer Science and Engineering (UBMK),

Taylor, R. E. (1999). A six-segment message strategy wheel. *Journal of advertising research*, *39*(6), 7-7.

Team, A. C. E. A. (2018). Alibaba Luban: AI-based Graphic Design Tool [Blog]. *Alibaba Cloud*. https://www.alibabacloud.com/blog/alibaba-luban-ai-based-graphic-design-tool_594294

Vakratsas, D., & Wang, X. S. (2020). Artificial Intelligence in Advertising Creativity. *Journal of Advertising*, *50*(1), XXX. https://doi.org/https://doi.org/10.1080/00913367.2020.1843090

van Noort, G., Himelboim, I., Martin, J., & Collinger, T. (2020). Introducing a Model of Automated Brand-Generated Content in an Era of Computational Advertising. *Journal of Advertising*, *49*(4), 411-427. https://doi.org/10.1080/00913367.2020.1795954

Wang, M., Yang, G.-W., Hu, S.-M., Yau, S.-T., & Shamir, A. (2019). Write-a-video: computational video montage from themed text. *ACM Trans. Graph.*, *38*(6), 177:171-177:113.

Wei, R., & Jiang, J. (2005). Exploring culture's influence on standardization dynamics of creative strategy and execution in international advertising. *Journalism & Mass Communication Quarterly*, *82*(4), 838-856.

West, D., Koslow, S., & Kilgour, M. (2019). Future directions for advertising creativity research. *Journal of Advertising*, *48*(1), 102-114.

Yang, W. (2019). *The effects of user-AI co-creation on UI design tasks*

Zhang, Y., Hu, K., Ren, P., Yang, C., Xu, W., & Hua, X.-S. (2017). Layout style modeling for automating banner design. Proceedings of the on Thematic Workshops of ACM Multimedia 2017,

Table 1. Creative Modification

Function	Application	Literature
Image colorization	https://deepai.org/machine-learning-model/colorizer The image colorization API may be used to colorize black and white photos or movies. Colorize old family pictures and historic images or use colorization to bring an old film back to life. https://github.com/pfnet/PaintsChainer Paints Chainer is a chainer-based line drawing colorizer. You may colorize your sketch semi-automatically using CNN.	ColorGAN (Ardizzone et al., 2019)
Covert photo to art	https://deepart.io/ Any snapshot may be turned into a piece of art.	DeepArt (Mao et al., 2017)
Videos enhance	https://www.topazlabs.com/video-enhance-ai For video upscaling, denoising, deinterlacing, and restoration, it uses information from many frames to obtain high-end results.	CNN-based (Lv et al., 2018)
Automated video editor	https://runwayml.com/ Masking, VFX, color correction, generation, and compositing	Consistent video editing (Kasten et al., 2021)
Face swap	https://www.rosebud.ai/ Convert face to a completely different face. Also known as deepfake. https://hey.reface.ai/ Face-swapping app to a social platform for personalized content creation	Deepfake (Nguyen et al., 2019)
Virtual try on	https://www.algoface.ai/ Makeup AR-tist allows your customers to virtually try on color cosmetics from their phones or in-store without having to open any boxes.	SmartFashion (Liang & Lin, 2021), (Mohammadi & Kalhor, 2021)

Table 2. Creative Conversion

Function	Application	Literature
Audio Transcription	**https://get.otter.ai/** Automated audio transcription and summarize key points	Multilingual LibriSpeech (MLS) (Tanberk et al., 2021)
Image captioning	https://vision-explorer.allenai.org/image_captioning Generating captions based on image with top 5 labeling	Top-Down Attention LSTM (Anderson et al., 2018)
Video understanding	https://anyclip.com/ Extract video data automatically, categorizing it with readily searchable phrases and categories.	Multimodal Video Generative Pretraining (MV-GPT) (Hongsuck Seo et al., 2022)

Table 3. Creative Generation

Function	**Application**	**Literature**
Text to speech	https://www.getwoord.com/ Synthetic voices replicate human-like natural sounding speech	DeepVoice (Arık et al., 2017)
Text to image	https://deepai.org/machine-learning-model/text2img http://gaugan.org/gaugan2/ Creates an image from scratch from a text description	DALL-E (Ramesh et al., 2021) GauGAN2 (Li et al., 2020)
Text to video	https://www.rephrase.ai/ http://www.synthesia.ai/ One human presenter appears in the center of screen and read your script https://design.ai/, https://lumen5.com/ have video database, generate video based on text	Write-a-video (Wang et al., 2019)
Chatbot	https://chat.openai.com/chat Ask any questions and response mimic human	Generative Pre-trained Transformer 3 (GPT3) (Stiennon et al., 2020)
Keywords	https://deepai.org/machine-learning-model/text-tagging From a sample of text, it extracts the most relevant and distinctive terms. These terms can then be used to sort documents into categories. Keywords can be extracted from a block of text. It selects keywords from the text at a lower temperature. It will create related terms at a higher temperature, which might be useful for generating search indexes.	Generative Pre-trained Transformer 3 (GPT3) (Stiennon et al., 2020)
Product description generation	https://textcortex.com/ Create a product title and target segment depending on the description.	Multiple Pointer-Generator Network (MPGN) (Hao et al., 2021)
Generated portraits	https://generated.photos/ Generate fake portraits and select different facial features.	style-based GAN architecture (StyleGAN) (Karras et al., 2019)

Table 4 Human and AI Involvement in Action

Key Function	Short Video Ads	Involvement of Humans	Involvement of AI	Image Ads	Involvement of Humans	Involvement of AI
Elements Center	Decompose video footage	None	AI algorithms	Generate design components	Human designers	None
Ad Maker	Produce short video ads	None	AI algorithms	Automated banner ad generation	None	AI algorithms
Creative Data Engine	Collect and analyze viewer responses	None	AI algorithms	N/A	N/A	N/A
Creative Optimizer	Identify effective frames and scenes	Advertising professionals	AI algorithms	N/A	N/A	N/A
Banner Ad Design	N/A	N/A	N/A	Design creative templates	Human designers	AI algorithms
Creative Templates	N/A	N/A	N/A	Develop creative templates	Human designers	AI algorithms