

Research Article

Digital Content Creation: Designing AI Assisted Smart Tool for End User Media Production

Huda Hamidon^{1,*}, Salliza Md Radzi², Nurul Shamimi Athirah Norizal³, and Fatin Nur Zulkipli⁴

¹ Faculty of Information Science Universiti Teknologi MARA; huda685@uitm.edu.my

² Faculty of Information Science Universiti Teknologi MARA; salliza@uitm.edu.my

³ Faculty of Information Science Universiti Teknologi MARA; 2023680766@student.uitm.edu.my

⁴ Faculty of Information Science Universiti Teknologi MARA; fatinnurzul@uitm.edu.my

* Correspondence: huda685@uitm.edu.my; +60137716589.

Keywords:

AI Assisted Smart Tool

GenAI

Digital content creation

KIE API

Visora

Abstract: Generative Artificial Intelligence (GenAI) is a smart tool that has reshaped media production in the technology driven era. However, most media production processes still require the use of multiple separate applications, high subscription costs and certain technical skills. This paper proposes the development of Visora, an AI assisted smart tool that integrates image generation, voice synthesis, and video composition into one. The system is suitable for educators, content creators and businesses who want to create graphic and video content without a complicated system.



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

The growing demand for online content on platforms like YouTube, TikTok, Instagram and a range of educational websites has increased the need for professionally designed images and narrated video materials. There is a rising need for content creators, educators and businesses to create high quality media, quickly and regularly. Content creators are social media influencers with large audiences on social media platforms (Kim et al., 2025). They use generative AI tools for many reasons including to produce imaginative content, improve efficiency and personalization (Lyu et al., 2024; Nalawade & Kadam, 2024). In the field of education, beyond traditional teaching approaches educators may utilize social media as an alternative medium for delivering educational content. The use of platforms such as Tik Tok enables high quality e-learning content that promotes nano learning experiences (Khlaif & Salha, 2021). Within the business perspective, video media serves as a fundamental element in performing social media marketing (Karol & Norman, 2019). The strategy of short form content has also led to a culture that reshapes consumer behavior, favoring visually appealing and narrated video content (Limbachiya, 2025). All of these provide evidence that Generative Artificial Intelligence (GenAI) systems are becoming increasingly important in the production of media content.

Conventionally, the creation of digital materials and content require good technical skills in using sophisticated video and graphic design software with high creativity as well as high software costs. However, the current technology has bridged the gap through automation and creative processes in the management of code, text, music, and video synthesis (Jadhav et al., 2024). The development of large language models (LLMs) and smart application programming interfaces (APIs) are also enabling more intelligent video generation directly from ordinary text inputs (Anusha et al., 2025; Sreerag et al., 2026). As a result, digital content creation is no longer confined to professionals but has become accessible to individuals with diverse levels of technical expertise.

At present, there are many tools available for assisting users in the content creation process. However, current solutions require users to switch between multiple platforms for image generation, voice synthesis, and video editing, resulting in fragmented workflows, higher costs, and reduced efficiency (Flew, 2021). In addition, professional and AI-powered content creation tools often involve expensive subscription fees, limiting access for students, educators, and small businesses (Lister et al., 2020). The lack of an integrated AI-driven platform further hinders productivity and content creation among nonprofessional users (Jenkins et al., 2018).

Therefore, this paper aims to propose an AI-integrated smart tool for media production namely Visora. The system uses artificial intelligence tools to enable users to create an image from a text description, generate an AI voiceover from any written script and automatically assemble the two into a complete video within a web browser. Rather than relying on the traditional media creation process that involves tasks such as image editing, voice recording, and video editing in separate applications, Visora integrates the whole media creation process.

2. LITERATURE REVIEW

In 2024, a study was carried out by performing a dataset analysis of more than 4 million artworks from over 50,000 users and found text to image AI enhanced creative productivity by 25% (Zhou & Lee, 2024). In addition, artificial intelligence is considered as a creative partner by designers where it has improved human creativity (Tang et al., 2024). Even so, previous research highlights that GenAI content presents both privacy and ethical challenges. Privacy concerns involve risks of data misuse, unauthorized content sharing, and adversarial attacks while ethical issues include copyright violations, biased outputs, misinformation, deepfakes, and threats to intellectual property rights (Chu et al., 2025; Al-kfairy et al., 2024; Suri et al., 2024).

Generative Artificial Intelligence (GenAI) is a smart tool that has reshaped media production in the technology driven era. GenAI enables users to generate images, audio, text, and videos with minimal expertise, making content creation more accessible to a wider audience. Despite these advancements, existing solutions often remain fragmented, requiring users to navigate multiple platforms and subscriptions to complete a single media production workflow. Content creation tool fragmentation is one of the biggest barriers for nonprofessionals participating in the digital media ecosystem (Flew, 2021).

Video content with voiceovers, for instance, requires the usage of several different tools. First, users can utilize image generator tools like Midjourney or Leonardo AI and record a voiceover using a tool like ElevenLabs or Murf AI. The image and voiceover then will be combined through a video editing tool, which may be Adobe Premiere Pro or CapCut. All of the tools have different processes and at times they must pay for every individual step. The absence of integrated digital media production tools has been recognized as a major factor that decreases productivity and discourages non-professional creators from consistently producing content (Jenkins et al., 2018).

Furthermore, the reliance on multiple applications introduces additional financial barriers, as users are often required to purchase separate subscriptions to access essential content creation functionalities. For example, the software provided by Adobe Creative Cloud, including applications like Premiere Pro and After Effects, is based on a monthly subscription that makes it impossible for students and other individual users working with a low budget to purchase and use it. Similarly, platforms such as Pictory AI and Synthesia, that automate part of the video production process, limit their valuable functionality exclusively to subscription-based pricing. Thus, the financial barrier results in an imbalance in which only rich people or companies may create high-quality narrated videos (Lister et al., 2020). Table 1 shows the benchmark analysis of Canva AI, Leonardo AI, Moonads and Pictory AI.

Table 1. Benchmark analysis.

Criteria	Canva AI	Leonardo AI	Moonads	Pictory AI
System features	Design tools, limited AI image generation via Magic Media, template based content creation, and graphic editing. No text to speech (TTS) narration or automated video composition.	Advanced AI image generation from text prompts, canvas editing, image-to image transformation, and model fine-tuning. No voice narration or video composition features	AI-powered UGC and marketing video production, AI influencer content, and product video ads targeted at businesses. No AI image generation from prompts and no personal gallery or history log.	Video creation from scripts and stock media, AI text to-speech narration with preset voices, and video export. No AI image generation from text prompts.
Performance	Image generation in 10–25 seconds on cloud servers. No video or audio processing. Performance is consistent but restricted to design and image outputs only.	Image generation in 5–20 seconds on cloud servers. No video or audio generation capability. Fast and reliable for image tasks but limited to single output generation.	Video ad generation in 1–3 minutes on cloud servers. Performance optimized for marketing content at scale. Dependent on internet connectivity and active subscription	Video generation takes 2–5 minutes on cloud servers. Performance depends on cloud server availability and subscription tier.
Security	OAuth and SSO authentication, token-based sessions, cloud stored user data and media subject to Canva's terms of service and data policies.	OAuth and SSO authentication, token-based session management, cloud-based data storage on third-party servers subject to Leonardo AI's data retention policies.	Email-based authentication, token-based session management, cloud stored media and user data subject to Moonads' privacy policies and third party cloud infrastructure.	Email-based authentication, token-based session management, cloud stored user media subject to Pictory AI's privacy and data retention policies.
Usability	Template driven interface with low learning curve. Designed for general users and casual designers. Does not support end to-end video creation workflow. Mobile responsive.	Customisable workspace with advanced generation controls. Medium learning curve suited to intermediate and advanced users. Not designed for non-technical users. Mobile responsive.	Marketing oriented interface designed for business users and performance marketers. Medium learning curve. Mobile responsive.	Multi-step video creation workflow requiring 6–8 steps. Medium learning curve. Designed for content marketers rather than general or academic users. Mobile responsive.

Efficiency	Somewhat automated design process with AI-powered design features. No automation for media production. Premium subscription needed for complete AI features - reducing efficiency for cost-sensitive users.	Automated generation of image from text. Only image output is automated. Additional tools needed to produce any audio or video, making it less efficient.	Automated video production system for marketing videos, designed for commercial production at scale. Cost efficient for commercial use but only for business marketing content. Subscription based, making it less accessible to non-commercial users.	Semi automated video production workflow from script to video. Limited stock media resources. Has a subscription model with limited free features, not cost-effective for one-off use.
------------	---	---	--	--

The above benchmark analysis matrix demonstrates that existing applications provide only partial support for the content creation workflow. Users must rely on multiple platforms to generate images, synthesize voiceovers, and produce videos. The comparison evaluates each system across five key dimensions like features, performance, security, usability, and efficiency that are consistent with standard software benchmarking methodology. Therefore, there is a need for an integrated solution that combines these functionalities within a single platform which can fill the gap in the existing content creation tools.

3. METHODOLOGY

The software development of Visora is based on Agile methodology. Agile is an iterative and incremental software development approach emphasising flexibility, software delivery on time and adaptability to changes throughout the development. Unlike the common "waterfall" approaches, Agile breaks the development into iterative cycles called sprints where each iteration produces a potentially shippable product (Beck et al., 2001). In general, iterative approaches are preferred in software engineering as those that are suited for software systems that interact with third-party technologies because it is required to understand the behaviour of the integrated system through implementation and testing.

Agile was selected for the project development for three reasons. First, the system uses several third-party APIs, namely KIE for AI image generation and elevenLabs for text to speech via KIE API. As a result, the data formats and other technicalities of the APIs can be only determined after the development. It would have been impossible to adopt a planned approach to development such as Waterfall where the complete and stable set of requirements must be defined up front.

Second, the scope and design of individual modules were changing throughout the development as new challenges appeared, which required further refinement and adaptation of design plans. For example, the video generation pipeline involving KIE TTS synthesis, audio file manipulation, and FFmpeg server-side video creation involved several design iterations due to integration-related challenges. All those changes were incorporated successfully into each iteration cycle of the project as per Agile without disturbing the development schedule.

Finally, this method is suitable since the project requires continuous testing and evaluation and incremental delivery of individual features. In total, there were nine sprints during the development of Visora, lasting fourteen weeks. Figure 1 shows the agile sprint cycle for Visora System development.

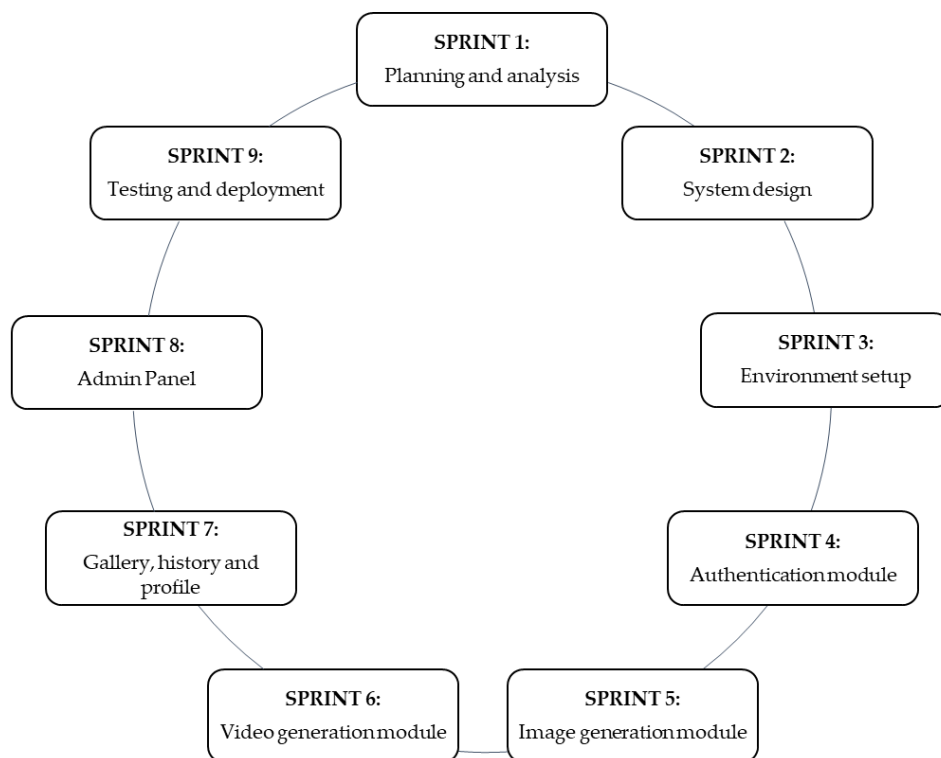


Figure 1. Visora System agile sprint cycle.

4. FINDINGS

The system innovation is called Visora, where it is an online system that offers an integrated, user-friendly and smart media production tool. The idea for Visora is based on the recognition that while AI systems for image generation and synthesised speech are now individually available, it's difficult to find a system that integrates these features in a manner that is feasible, cost-effective and user-friendly for the average person. Current systems either focus on a single type of media production or are highly technical. This presents an opportunity to create a system that makes AI-powered media creation more accessible.

Visora is a web-based system that uses HTML, CSS and JavaScript for the front-end, PHP for the back end and MySQL as the database management system using XAMPP. The application interacts with the KIE API for AI image generation and ElevenLabs text-to-speech as well as FFmpeg which composes server-side video. The app offers four key features which are accessible by tabbed navigation. For instance, Content Creation, Gallery, Profile and History that is designed to manage the entire life cycle of a user's creative process content generation and management.

The project includes user authentication, AI-based image generation from text, text-to-voice narration for any type of media such as stories, educational content, commentary and factual descriptions, video generation using the generated images and synthesised audio, and a media management system and gallery with activity history. The system is intended to be deployed locally via XAMPP and is therefore ideal for educational institutions, individuals and small groups looking for an affordable, hosted solution.

The relevance of Visora to real-world applications is significant. For educators, teachers and students can quickly create narrated visuals for educational content and presentations. In content production, YouTube vloggers and social media influencers can automate the creation of video with AI generated narration without the need for video editing expertise. In business, small enterprises can automatically create visual media with professional sounding voiceovers for marketing purposes. Visora is thus not only a technically robust project, but also a relevant response to an issue faced by many digital content creators.

4.1 Dashboard Overview

The application provides a platform that facilitates interactions between two entities which are user and admin. This part will discuss the main dashboard for both roles. Figure 2 shows the Admin Dashboard where it provides the overview of recent system activities along with the analysis or summary of total users based on the categories as well as the total images and videos generated. Visora admin are allowed to manage user accounts, manage all content, view system history, monitor activity logs and delete requests.

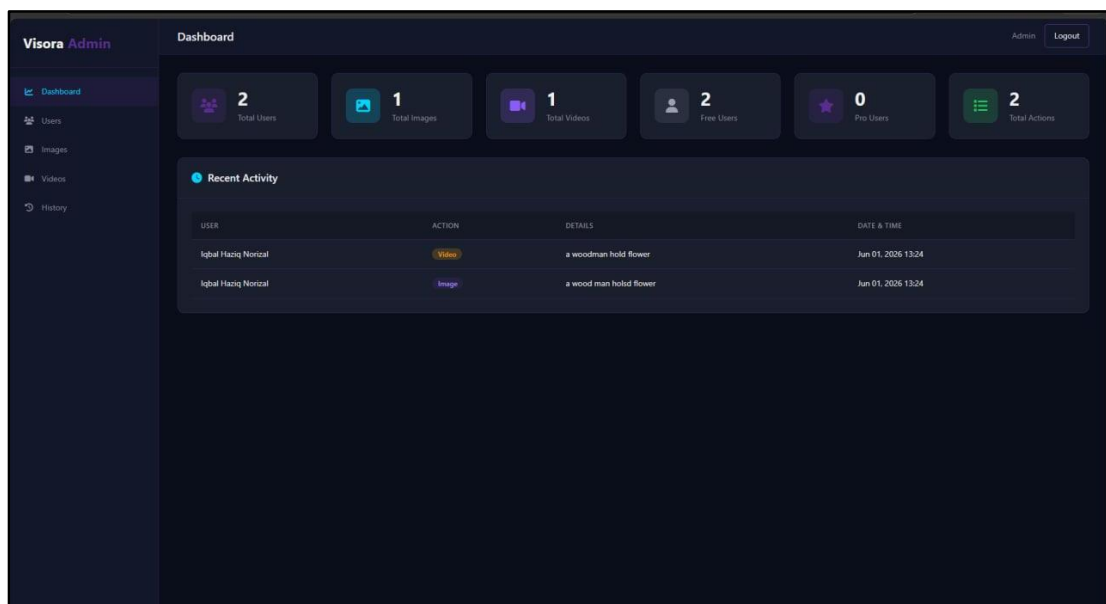


Figure 2. Admin Dashboard.

As for user roles, it can be categorized into two which is Free User and Pro User. Free users may generate 3 AI images and 1 AI video while for Pro Users, they can generate 10 AI images and 5 AI videos. Figure 3 shows the User Dashboard, a page where users can create new media using prompt. All the generated media will be stored in the personal gallery as shown in Figure 4.

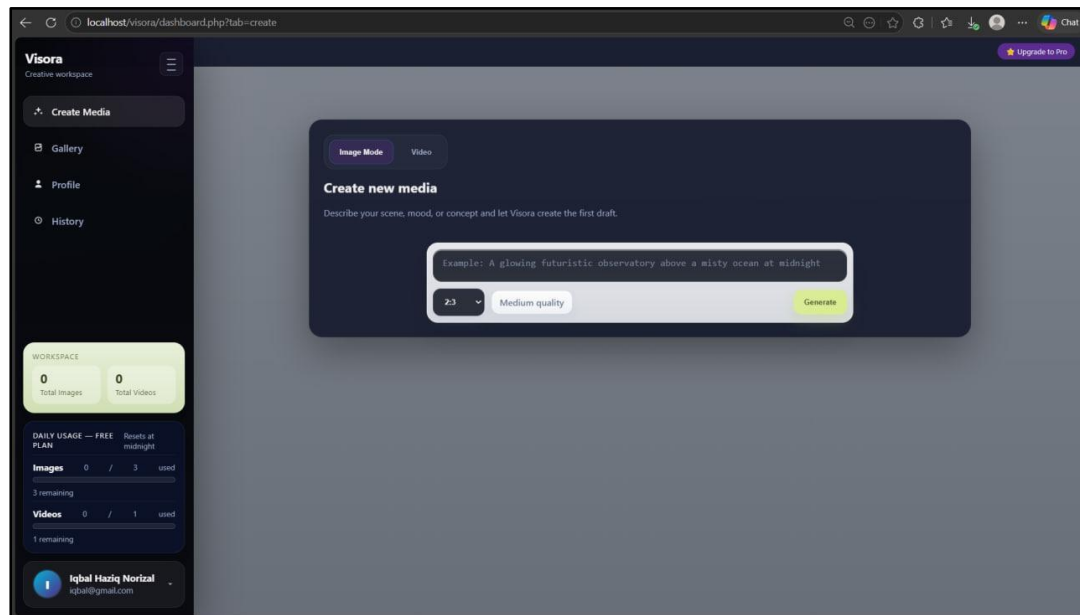


Figure 3. User Dashboard.

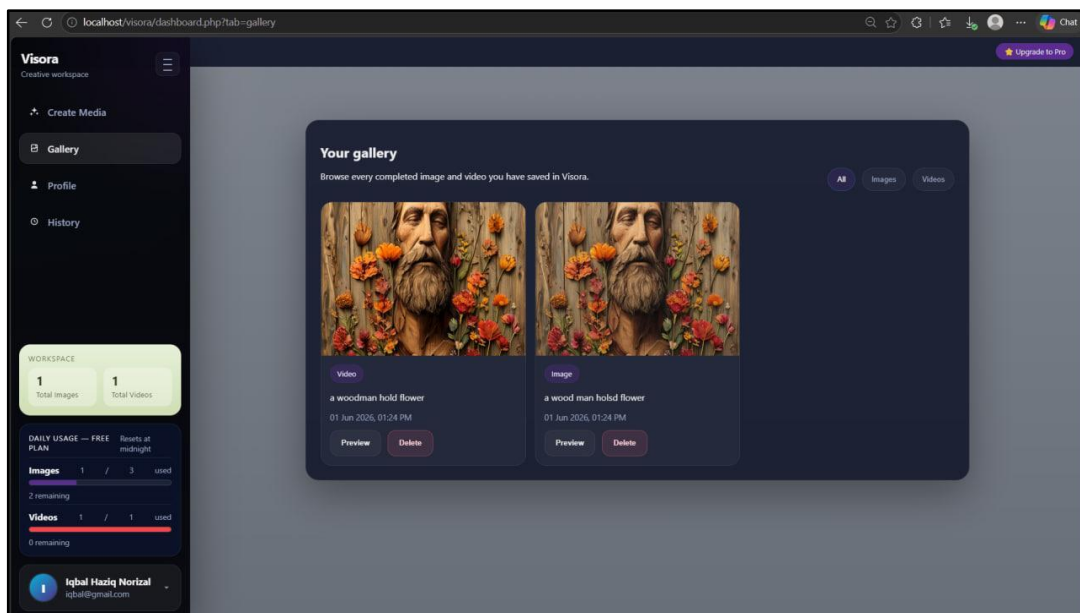


Figure 4. User Gallery.

5. DISCUSSION

The development of the system greatly impacted content creators, educators and businesses. Visora provides a combination of features and access that is unique among existing video platforms. It offers a unified system that integrates AI image generation, text-to-speech voice narration, and video composition, together with a personal media library, an activity log, and user account management. In terms of speed, Visora completes the entire image-to-video processing pipeline in less than 90 seconds on a local server.

The application also considered the security aspect by storing all user media in a local XAMPP server and preventing user media from being exposed to third-party cloud storage and data retention policies. In terms of user-friendliness, Visora has the shortest workflow for completing a video production process and is the only system built for non-technical users such as students and teachers. Finally, in terms of efficiency, Visora is the only platform that provides an automated end-to-end content creation pipeline that can be deployed for free in a local environment, enabling its use by educational institutions and individual creators with no financial barriers. These results confirm that Visora fills an important gap in the existing content creation tool ecosystem.

6. CONCLUSION

In conclusion, Visora successfully addresses three core challenges faced by digital content creators today which are fragmented tools, expensive fee and absence of unified solution. The system solve the issues by offering a web-based system built with HTML, CSS, JavaScript, PHP, and MySQL (via XAMPP) that enables any user to generate an AI image based on a text description using the KIE API, input any text content as the narration script, create a cinematic AI voiceover narration using the ElevenLabs text to-speech service through the KIE API, and automatically produce a completed MP4 video using FFmpeg, all within a single tabbed interface in the application itself. The narration content input, which can be any content that a user wishes to create like stories, educational content, commentary, facts and descriptions that makes Visora an accessible platform for a wide range of content creation scenarios rather than limiting the user to one content format.

Visora not only offers a functional and feature rich, but also accessible, secure and free to use system that represents a valuable contribution to both the research areas of information management systems and AI-powered media production systems. There is a high potential for commercialization due to the growing demand for accessible and affordable GenAI technologies.

Acknowledgments: Appreciation goes to the Faculty of Information Science and Universiti Teknologi MARA for providing the resources in making this project possible. Warmest thanks to family members and friends for all the motivation and inspiration.

References

- Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *Informatics*, 11(3), 58. <https://doi.org/10.3390/informatics11030058>
- Anusha, S., Harini, R., Nikita, T., Priyanka, V. N., Priya, G. N., & Babu, P. V. (2025). VISOGEN: AI and APIs unleashed for smarter video generation. *Proceedings - 2025 7th International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 467-473. <https://doi.org/10.1109/CCICT65753.2025.00077>
- Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., & Thomas, D. (2001). *Manifesto for Agile Software Development*. Retrieved from agilemanifesto.org
- Chu, C., Zheng, Y., Lu, J., Zeng, T., & Li, T. (2025). P-3.1: Navigating privacy concerns: A comprehensive review of challenges in artificial intelligence image-to-video generation applications. *SID Symposium Digest of Technical Papers*, 56(S1), 864–868. <https://doi.org/10.1002/sdtp.18951>

- Flew, T. (2021). *The creative industries: Culture and policy* (2nd ed.). SAGE Publications. <https://doi.org/10.4135/9781526484406>
- Jadhav, S., Patil, J. A., Rachappa Jyoti, R., Kanase, O. S., & Satish Khadde, M. (2024). The future of content creation: Leveraging AI for code, text, music, and video generation. *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. <https://doi.org/10.1109/ICCUBEA61740.2024.10775258>
- Jenkins, H., Ford, S., & Green, J. (2018). *Spreadable media: Creating value and meaning in a networked culture*. New York University Press. <https://doi.org/10.18574/nyu/9780814743508.001.0001>
- Karol, K., & Norman, C. (2019). The rise of, and need for, video in social media marketing. *Journal of Education Advancement & Marketing*, 3(4), 316–324. <https://doi.org/10.69554/PCKH5100>
- Khlaif, Z. N., & Salha, S. (2021). Using TikTok in education: A form of micro-learning or nano-learning? *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 12(3), 213–218. <https://doi.org/10.30476/ijvlms.2021.90211.1087>
- Kim, J., Yu, M., Zhi, J., Milani, S., Cheng, J., & Bernstein, M. S. (2025). Content creation with generative AI: How do content creators responsibly use generative AI tools? *Proceedings of the ACM on Human-Computer Interaction*, 9(CSCW1), Article 9. <https://doi.org/10.1145/3788080>
- Limbachiya, M. A. (2025). The digital shift: How short-form content and instant gratification are reshaping consumer behaviour and patience. *International Journal of Scientific Research in Engineering and Management*, 9(5), 1–9. <https://doi.org/10.55041/IJSREM46304>
- Lister, M., Dovey, J., Giddings, S., Grant, I., & Kelly, K. (2020). *New media: A critical introduction* (3rd ed.). Routledge.
- Nalawade, S. R., & Kadam, A. S. (2024). Artificial intelligence in content creation: Technologies, applications, and challenges. *Journal of Information, Decision Sciences and Big Data Management*, 1(2), 22–30.
- Sreerag, A., P, M., A, A., Ranjit, A., & Babu, A. S. (2026). Multimodal media creation: Integrating LLMs for high-quality video generation. *Lecture Notes in Networks and Systems*, 1656, 167-177. https://doi.org/10.1007/978-3-032-06703-6_16
- Suri, C. S., Shukla, S. S., Pal, R., & Kumar, M. B. (2024). Generative AI in content creation: Opportunities and ethical challenges. In *2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE. <https://doi.org/10.1109/UPCON62832.2024.10983285>
- Tang, Y., Ciancia, M., Wang, Z., & Gao, Z. (2024). *What's next? Exploring utilization, challenges, and future directions of AI-generated image tools in graphic design* (arXiv No. 2406.13436). arXiv. <https://doi.org/10.48550/arXiv.2406.13436>
- Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, 3(3), pgae052. <https://doi.org/10.1093/pnasnexus/pgae052>