International Journal of Economics, Commerce and Management
United Kingdom ISSN 2348 0386 Vol. 13, Issue 4, April 2025

https://ijecm.co.uk/

BEYOND AUTOMATION: A QUALITATIVE STUDY ON THE REMAINING CHALLENGES FOR GENERATIVE AI IN ADVERTISING DESIGN

Shueh-Cheng Hu
Providence University, Faculty of Computer Science and Communication Engineering, Taiwan
shuehcheng@gmail.com

Abstract

Nowadays, Generative AI (GAI) tools have become a valuable assistant in the workplace, gradually transforming the way we work. However, many users have reported a decline in user experience when using GAI tools, leading to abandonment or switching to alternative solutions. This study uses a qualitative research method: critical incident technique (CIT) to reveal the limits of GAI tools on advertising design. The study aims to investigate factors that negatively impact user experience. The research findings reveal eight factors - originality, creativity, logical coherence, depth, flexibility, emotional resonance, homogeneity, and precision, they collectively contribute to diminished user experience. In the end, this research work provides actionable suggestions for those who need an improved method of planning and designing advertisements by using GAI tools, which ultimately enhance user experience.

Keywords: Generative AI, Critical Incident Technique, Advertising Design, User Experience

INTRODUCTION

In the context of rapid advancements in science, technology, and continuous innovation in data analysis, artificial intelligence has deeply permeated our daily lives. The ability of AIGC technology to quickly generate text based on user requirements has significantly increased writing efficiency, making it an indispensable assistant in the workplace. The remarkable debut of ChatGPT in late 2022 positioned it as one of the most prominent representatives of AI technologies. In the following years, people across various fields have leveraged different

generative AI tools' capabilities to support their professional works, and advertising designers are no exception. Despite their exceptional performance in areas such as knowledge integration, thought mirroring, multimodal responses, and real-time information bridging, GAI tools still exhibits certain shortcomings in the overall process of advertising planning and design.

This study utilizes online questionnaires and employs the critical incident technique (CIT) to explore the negative perceptions of users while they were using GAI tools for advertising design. The work aims to provide valuable suggestions for improving GAI tools and enhancing user experience to better serve advertising design needs.

LITERATURE REVIEW

Artificial Intelligence-Generated Content (AIGC) refers to technology that automates the creation, editing, and optimization of diverse digital content through algorithmic models, with the core aim of enhancing content production efficiency and accessibility. In recent years, breakthroughs in generative AI have extended AIGC from single-modality text or image generation to multimodal interactions, enabling cross-modal collaborative content creation that facilitates comprehensive connections in advertising planning (Cao et al., 2025). On the technical front, the rise of mobile edge networks has propelled AIGC services toward real-time and personalized solutions, allowing users to swiftly obtain tailored advertising plans while maintaining privacy (Xu et al., 2024). Meanwhile, the application of AIGC in the commercial sector has deepened, with AIGC software providing automated feedback based on user demands and enhancing user interaction through inquiry mechanisms. A literature survey provided a wide range of concrete samples of how GAI tools have been applied in advertising (Ercan et al., 2025). However, challenges such as user trust issues regarding generated content remain unresolved (Jiang et al., 2024). In addition, AIGC's penetration into the design field has triggered industry transformations. AIGC software aids in fostering creative divergence to complete planning proposals and leverages feedback data to make further strategic decisions (Wu et al., 2024). Nevertheless, global AIGC research demonstrates significant imbalances, with technological advancements heavily concentrated in economically developed regions and insufficient interdisciplinary integration. This indicates an urgent need to optimize resource allocation and strengthen international cooperation to drive progress across industries (Guo et al., 2025). In summary, AIGC is gradually becoming a driving force for innovation across multiple industries. However, challenges in areas such as innovation, privacy protection, and technological accessibility still require further improvement.

Beyond discussion of AI tools' capability or functionality, there is a research work focus on the ethical issues regarding the usage of GAI tools (Berrah et al., 2024). Although the open-source GAI tools reduces labor costs for various enterprises, it still requires further optimization and development to address its imperfections (Parghi et al., 2025)

RESEARCH METHODOLOGY

Critical Incident Technique (CIT)

The CIT was introduced by the American scholar Flanagan in 1954 as a qualitative research method designed to delve into human behavior, performance, or psychological phenomena. Initially applied in military psychological studies due to its focus on critical incidents, CIT has since been widely adopted across various fields.

As a well-established qualitative research method, CIT has demonstrated its value in systematic analysis and practical guidance across multiple domains. In the field of education, CIT has been used to analyze levels of teacher reflection, revealing the influence of individual factors (such as teaching experience and educational background) on the depth of reflection (Voulgari & Koutrouba, 2024). In healthcare, CIT has identified dynamic factors of clinical risk, such as unplanned extubation incidents caused by environmental chaos or communication failures, thereby facilitating the formulation of targeted intervention strategies (Danielis et al., 2018). In the business domain, the combination of CIT with quantitative methods has overcome the linear assumptions of service quality, verifying the nonlinear impact of low-frequency events (such as complaint responses) on customer satisfaction (Tontini et al., 2017). In tourism management, CIT has been employed in studies on zoo management and supply chain education to investigate and interpret boundaries and interdisciplinary adaptability (Rashid-Radha et al., 2021; Schulze et al., 2021). In summary, CIT has been widely applied in research across multiple fields with significant success, earning broad recognition in academia. Both its maturity and reliability demonstrate its high level of utility. Based on this, the present study utilizes CIT to thoroughly analyze the critical negative incidents experienced by users when employing GAI tools for advertising design. The aim is to identify factors that may reduce user experience and provide actionable recommendations for improvement.

Research Design

Given that users often employ GAI tools multiple times to assist with advertising design, it is essential to delve into their in-depth experiences during this process. Therefore, this study adopts the qualitative research method—Critical Incident Technique (CIT). While most users express overall satisfaction with GAI tools' performance, there are areas that require

improvement. To further advance GAI tools and similar AIGC models, this research focuses specifically on advertising design, collecting only users' negative critical incidents experienced while using GAI tools (Flanagan, 1954) and designs the questionnaire based on the method proposed by Bitner et al. (1990).

The Data

The data using questionnaire were collected over a period of 14 days, from March 3 to March 16, 2025. After systematic organization and thorough analysis, the data provided essential insights into reducing the negative experiences associated with using GAI tools for advertising design. The questionnaires were gathered online and included a screening criterion: "Have you used GAI tools for advertising design three or more times and applied it to practical advertising design tasks?" This ensured the validity of the collected data. Based on this, further responses were collected from the screened participants regarding the critical negative incidents they encountered while using GAI tools, alongside their suggested improvements and willingness to continually use the tool.

ANALYSIS AND RESULTS

Basic Data

A total of 436 questionnaires were collected for this study. After removing 15 questionnaires that were off-topic or irrelevant, 421 valid questionnaires were obtained, resulting in 421 critical negative incidents for further data analysis. Flanagan suggests that while thousands of critical incidents may be required for studying complex activities, but 50 to 100 incidents are sufficient for investigating relatively simple activities. With 421 critical incidents included in the analysis, the sample size meets the requirements of the Critical Incident Technique. The basic demographic information of the respondents is shown in Table 1.

Table 1: Basic Information of Respondents

Variable	Measurement Item	Sample Percentage
Gender	Male	53.5%
	Female	46.5%
Age	18 and below	8.3%
	19-24	29.4%
	25-30	32.7%
	31-40	19.0%
	41-55	8.5%
	55 and above	3.1%
Education Level	High school or below	17.1%
	College (& Associate's)	33.0%
	Bachelor's degree	37.1%

	Graduate and above	12.8%
Monthly Income	≤ 3500	15.2%
	3501-6000	31.8%
	6001-8500	23.1%
	8501-12000	17.1%
	≥ 12001	12.8%
Employment Status	Student	18.0%
	Employed	54.5%
	Unemployed/Job Seeking	17.3%
	Retired	10.2%

Table 1…

Classification Principles

In this study, a total of 421 critical negative incidents were collected. After preliminary review by the researchers, these incidents were categorized into eight classifications: originality, creativity, logical coherence, depth, flexibility, emotional resonance, homogeneity, and precision. The names and detailed descriptions of these categories can be found in Table 2.

Table 2: Naming and Explanation of Key Events

Category Name	Detailed Description
Originality	Refers to the autonomy of GAI tools in generating advertising contents, including whether the content is merely copied and pasted from other works.
Innovation	Refers to the novelty of GAI tool-generated contents, including the use of technology, filming angles, color application, storyline, etc.
Logic	Refers to the rationality of the causal chain between the advertising contents and execution, including the internal storyline connections and the coherence between characters and dialogues.
Profundity	Refers to GAI tools' comprehensive understanding of the designed contents, including in-depth interpretations of ethnic cultures and multi-layered meanings in literary works.
Flexibility	Refers to the adjustability of GAI tools designed contents based on market, season, location, audience, weather, and other dynamic factors.
Emotional Appeal	Refers to the extent to which GAI tools generated contents resonates with users through emotional experiences, including sensory design, emotional curve design, and storyline development.
Homogeneity	Refers to the degree of variation in GAI tools generated contents, including differences in storylines, filming angles, and dialogue usages.
Accuracy	Refers to GAI tools' responsiveness to user needs and the precision of the information conveyed in the feedback contents.

To ensure the reliability and usability of the data in this study, experienced experts in advertising design were invited to serve as classifiers for the collected critical incidents. After the researchers initially reviewed and categorized the incidents, the classifiers further refined and categorized the data. To guarantee the accuracy of the categorization and the advancement of

the research, the classifiers reclassified the critical incidents 30 days later. Relevant information regarding the classifiers is presented in Table 3.

Table 3: Classifier Background Information

Classifier	Background Information
Classifier 1	A university lecturer specializing in online advertising, with nearly 10 years of experience in advertising design courses and closely monitoring changes in the advertising market.
Classifier 2	An advertising planner at an advertising company, deeply engaged in the advertising market for 6 years with extensive practical experience in advertising planning and design.
Classifier 3	A copywriter at an advertising company, specializing in advertising copywriting for over 5 years, with rich experience in crafting advertising contents.

Reliability and Validity Analysis

Reliability

Reliability analysis is a method used to assess the consistency and stability of research instrument. Within the overall framework of the Critical Incident Technique (CIT), reliability testing evaluates both inter-classifier consistency and individual classifier consistency. Inter-classifier consistency ensures the methodological coherence of the study, while individual classifier consistency assesses methodological stability, ensuring that the research approach aligns with the intended objectives (Flanagan, 1954). Data indicate that when reliability test results exceed 0.8, the data can be considered consistent and stable, thus suitable for further research (Butterfield et al., 2005). Details on key event classifier consistency are shown in Table 4.

Table 4: Classifier Consistency

Classifier	Classifier1	Classifier2	Classifier3
Classifier1	366	-	-
Classifier2	307	392	-
Classifier3	286	324	349

Using the data in Tables 4 and 5 to validate the reliability analysis of the 3 classifiers' categorization. The following formulas were used to calculate the reliability of the classifiers' categorization:

$$A = \frac{\frac{2M_{12}}{n_1+n_2}+\frac{2M_{23}}{n_2+n_3}+\frac{2M_{13}}{n_1+n_3}}{N}$$

$$R = \frac{(N\times A)}{1+[(N-1)\times A]}$$

Where, R represents reliability. A is the average level of inter-coder agreement. N is the number of classifiers. M denotes the number of identical samples categorized by different classifiers

(e.g.M_{12} represents the number of identical samples categorized by classifier 1 and classifier 2). n indicates the number of identical samples categorized by the same classifier in two separate instances (e.g.,n_{12} represents the number of identical samples categorized by classifier 1 in two instances). Based on the formulas above, calculation results are shown in Table 5.

Table 5: Classification Reliability Table

BBT Classification	Average Mutual Consistency (A)	Reliability (R)
Unsatisfied	0.828	0.935

The data indicate that both the average inter-coder agreement and reliability exceed 0.8. This means that the categorization labels used in this study effectively summarize all critical incidents, with clear boundaries that accurately reflect the characteristics of the critical incidents. Therefore, the reliability test is successful, ensuring that the study provides authentic and valid support for subsequent research.

Validity

Validity analysis evaluates whether the research methods accurately reflect the true attributes of the intended construct. It primarily includes expert validity, content validity, and construct validity (Haynes et al., 1995). Expert validity refers to the professional judgment of experts in relevant fields to ensure that the content of the research methods aligns closely with the conceptual framework and practical needs of the research objectives (Berk, 1990). In this study, three advertising planning and design experts were invited to classify the critical incidents to ensure that the categorization data remained consistent with the characteristics of advertising design, meeting the requirements of expert validity. Content validity measures whether the questionnaire items in the research method comprehensively and appropriately represent the relevant dimensions of the intended construct (Lawshe, 1975). The questionnaire in this study includes items such as advertising slogans, dynamic headlines, video scripts, multi-panel comic creation, storyboard generation, and long-form copywriting, which cover relevant dimensions of advertising design, meeting the requirements for content validity. Finally, construct validity assesses whether the research methods align with their intended conceptual framework through dynamic interaction between theoretical and empirical data (Cronbach & Meehl, 1995). This study employs qualitative CIT research methods to deeply investigate users' negative experiences when using GAI tools on advertising design. The collected data consist entirely of users' negative critical incidents, meeting the requirements for construct validity.

Categorization Results

To gain a deeper understanding of the impact of each category of negative critical incidents on user experience, two critical incidents from each category were selected as examples after the incidents were classified and analyzed. Examples of these critical incidents are shown in Table 6.

Table 6: Key Event Examples

Event Category	Example 1	Example 2
Originality	The generated ad copy is not original, as it plagiarizes existing content from other brands, making it prone to infringement.	Formulaic content is widespread, lacking originality. AI models rely on frequently used words and popular phrases in training, no uniqueness shown.
Innovation	The creative direction is too narrow. GAI tools tends to repeat specific creative directions or word combinations when generating dynamic titles, resulting in insufficient title diversity.	Lack of creativity. Ad storyboards require strong creative and visual expression, but GAI tools-generated storyboards are dull, lacking distinct creativity and visual impact.
Logic	The generated storyline is disjointed and lacks coherence in sequencing shots.	The ad dialogues designed for me are completely opposite to how people actually speak, even appearing frightening, making them completely illogical.
Profundity	The ad is intended to promote ethnic culture, but AI tools' generated contents remain superficial, failing to deeply showcase ethnic characteristics.	The expressions in my advertising plan remain at a surface level, without delving into what users truly want.
Flexibility	The generated advertising contents cannot be quickly adjusted for specific target audiences within a short time; I must repeatedly modify the settings.	The designed contents do not adapt to different application scenarios and only adjusts in one limited direction.
Emotional Appeal	Insufficient emotional resonance. For instance, in titles that need to convey warmth, sentiment, or passion, but GAI tools struggles to accurately capture and express these emotions.	The AI generated contents does not effectively enhance emotions such as a father’s guilt or a daughter’s emotional reaction through language or imagery.
Homogeneity	The material library is limited, and templates are overly standardized, making the generated plans appear very close to others.	When I set three different requirements, but the results were nearly identical, as if they were generated using the same mold.
Accuracy	AI tools does not understand my requests precisely. I need to refine them 3-5 times before it generates contents that aligns with my expectations.	I clearly stated that my target audience is teenagers, yet GAI tools only recognized this after several reminders.

After categorizing the critical incidents, this study conducted further statistical analysis of the sampled data. Detailed data are presented in Table 7.

Table 7: Classification Data Ranking and Degree Analysis

Classification Name	Frequency	Percentage of all incidents
Logic	84 cases	19.95%
Accuracy	77 cases	18.29%
Profundity	63 cases	14.96%
Innovation	59 cases	14.01%
Emotional Appeal	45 cases	10.69%
Homogeneity	45 cases	10.69%
Flexibility	32 cases	7.60%
Originality	16 cases	3.80%

From the above data, it can be observed that critical incidents related to logical coherence, creativity, and depth account for a significant proportion. This indicates that users place great importance on the logical structure, precision, and depth of AI generated contents when using GAI tools in advertising design. Therefore, focusing on optimizing these three aspects can maximize user satisfaction with the overall experience.

CONCLUDING REMARKS

Conclusion

This study applied the Critical Incident Technique (CIT) to focus on the use of GAI tools in assisting advertising planning and explored the factors that diminish user experience. The research findings revealed that eight dimensions—originality, creativity, logical coherence, depth, flexibility, emotional resonance, homogeneity, and precision—impact user experience and reduce satisfaction levels. Among these, logical coherence, precision, and depth have the most significant effect on user experience.

Recommendations

Although research on GAI tools and AIGC technologies is steadily increasing, studies that specifically focus on advertising design as a lens to explore factors that reduce user experience remain relatively scarce. This study conducted an in-depth analysis of the factors negatively influencing user experience. Based on the research findings of the present work, two groups of suggestions are proposed accordingly.

Suggestions for GAI tools developers

GAI tools' researchers and developers must strengthen the analysis of factors that reduce user experience in order to achieve better user favorability and engagement. The following four suggestions are provided for advertising design: first, to guide users in expressing their needs. Although current AIGC software is simple to use, it cannot ensure that every user can accurately express their needs every time. In this case, AIGC software needs to use inquiry mechanisms to understand user requirements to guarantee comprehensive and appropriate answers. Second, continuously optimize the content library. Although GAI tools currently have highly diverse contents, users may experience aesthetic fatigue over time. Therefore, GAI tools should regularly update and optimize their internal content to meet users' evolving needs. Third, optimize algorithms and strengthen filtering capabilities. Users may encounter inaccurate information when using GAI tools for advertising design. It is recommended that GAI tools can enhance active detection technologies in order to ensure the accuracy of feedback information. Fourth, focus on social impact and increase information transparency. At present, the public's attitude towards AI remains largely skeptical, and further efforts are required to win users' trust and foster engagement. Therefore, when users interact with GAI tools, they should clearly disclose the probabilistic characteristics of generated content, its sources, and associated risk warnings to increase user trust.

Suggestions for GAI users who design advertisements

The initial purpose of artificial intelligence was to help humanity achieve better development; however, we must remain cautious during its usage. This study proposes four suggestions for users: first, select GAI tools that suit your needs. Since each tool has its specialized application areas, users should clarify the specific aspects where AI assistance is required and choose AI tools that best meet those needs. As such, users must carefully screen and select tools that both meet their requirements and demonstrate high efficiency. Second, continuously learn and adapt to GAI tools. GAI tools are constantly advancing, and to keep pace with societal progress and improve work efficiency, users need to continuously develop their skills in using these tools, eventually achieving proficiency. Third, maintain critical thinking. The contents generated by GAI tools originally come from Web contents in large, so inaccuracies and hallucination in generated content is inevitable. Besides, to advertising designers who demand a high level of originality, any instances of plagiarism in advertising contents might lead to severe consequences. Therefore, users must always maintain critical thinking; verify the contents generated by AI tools, rather than blind trust. Fourth, safeguard personal privacy and data security. Although network security has significantly improved, risks still exist. When using

GAI tools, users should avoid disclosing personal privacy, trade secrets, or other sensitive information. In addition, they should pay close attention to the data usage policies of AI tool providers. Priority should be given to GAI tools with encrypted transmission and anonymization features.

Suggestions for the Future Research

The present work could be extended toward two facets: first, because GAI tools could be used in creating contents for different application domains, such as e-learning, e-entertainment, etc. Thus, the similar qualitative study could be conducted to reveal challenges that people will encounter while using GAI tools to create contents in different domains. The second facet involves the research works using quantitative analysis alongside a qualitative approach to measure content generated by GAI tools based on numerical criteria such as satisfaction ratings and content accuracy.

REFERENCES

Berk, R. A. (1990). Importance of expert judgment in content-related validity evidence. Western Journal of Nursing Research, 12(5), 659–671.

Berrah, L., Trentesaux, D., & Guerre-Chaley, F. (2024, September). Ethical issues in the use of Generative Artificial Intelligence in performance management: Industrial case studies. In SOHOMA 2024.

Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. Journal of Marketing, 54(1), 71–84.

Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A. S. T. (2005). Fifty years of the critical incident technique: 1954–2004 and beyond. Qualitative Research, 5(4), 475–497.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P., & Sun, L. (2025). A survey of AI-generated content (AIGC). ACM Computing Surveys, 57(5), 1–38.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281.

Ercan, H. D., Tanrıverdi, N. S., & Taşkın, N. (2024). A Systematic literature review for Artificial Intelligence in Advertising.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51(4), 327–358.

Guo, Y., Yu, F., Lai, J., & Yuan, X. (2025). How is AIGC shaping the world: An analysis of bibliometrics. Information Research: An International Electronic Journal, 30(iConf), 679–689.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment, 7(3), 238–247.

Jiang, X., Wu, Z., & Yu, F. (2024). Constructing consumer trust through artificial intelligence-generated content. Academic Journal of Business & Management, 6(8), 263–272.

Lawshe, C. H. (1975). A quantitative approach to content validity. Personnel Psychology, 28(4).

Parghi, N. B., Chauhan, C. R., & Karavadiya, D. A. (2025). A comparative study of DeepSeek and other AI tools. International Journal of Innovative Science and Research Technology, 10(3), 1125–1132.

Rashid-Radha, J. Z., Shariff, N. M., Kasim, A., Ghazali, R. M., & Mohamed, A. E. (2021). Exploring how zoo servicescapes impact memorable tourism experience: A critical incident technique (CIT) approach. Journal of Tourism, Hospitality and Environment Management, 6(23), 53–59.

Schulze, H., Bals, L., & Warwick, J. (2021). Teaching sustainable purchasing and supply management competences using critical incidents. Operations Management Education Review, 15, 201–222.

Tontini, G., dos Santos Bento, G., Milbratz, T. C., Volles, B. K., & Ferrari, D. (2017). The critical incident technique (CIT) and penalty-reward contrast analysis (PRCA) applied to online reviews on TripAdvisor: Evaluation of satisfaction of hotel customers. International Journal of Hospitality Management, 66(1), 106–116.

Voulgari, R., & Koutrouba, K. (2024). Examining the depth of primary school teachers' reflection through the critical incident technique. Educational Studies, 50(5), 640–658.

Wu, J., Cai, Y., Sun, T., Ma, K., & Lu, C. (2024). Integrating AIGC with design: Dependence, application, and evolution—a systematic literature review. Journal of Engineering Design, 1–39.

Xu, M., Du, H., Niyato, D., Kang, J., Xiong, Z., Mao, S., & Poor, H. V. (2024). Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services. IEEE Communications Surveys & Tutorials, 26(2), 1127–1170.