CURRENT ISSUES OF ECONOMIC SCIENCES

Marketing

UDC 004.8:004.93:659.1

DOI https://doi.org/10.5281/zenodo.20116037

Automation of creative hypothesis testing in video marketing based on large language models and computer vision

Mariia Hladkova,
Master, Senior Videographer and Video Editor,
Golden Pet Brands, El Segundo, USA
https://orcid.org/0009-0005-3966-2540

Accepted: 12.03.2026 | Published: 30.03.2026

Abstract. The relevance of the study is determined by the growing role of video marketing in the digital environment, the increasing complexity of audiovisual content structures, and the rising requirements for the validity of managerial decision-making, which limits the effectiveness of traditional approaches to testing creative solutions and actualizes the need for implementing automated analytical systems. **The purpose** of the study is defined as the development of an integrated methodological framework for automating the testing of creative hypotheses in video marketing based on large language models and computer vision in order to enhance the validity, efficiency, and scalability of data-driven decision-making processes. **Methods.** The study applies methods of analysis and synthesis to generalize theoretical approaches to the automation of creative hypothesis testing; systematization – to structure their components and functional characteristics; comparative analysis – to assess the capabilities of large language models and computer vision; and logical generalization – to develop methodological approaches and practical recommendations. **Results**. The essence of creative hypotheses as multidimensional analytical constructs integrating semantic, visual, and behavioral parameters of content has been investigated. It has been revealed that the use of multimodal technologies ensures a transition to continuous testing and adaptive optimization of video content. It has been proven that the integration of large language models and computer vision into a unified analytical system increases the accuracy of content effectiveness evaluation, reduces experimentation costs, and minimizes the time lag between content creation and analysis. It has been established that the effectiveness of automation is constrained by issues of model interpretability, data quality and consistency, as well as the limitations of performance evaluation systems. **Conclusions.** It has been substantiated that improving the effectiveness of automated testing is achieved through the formation of an integrated multimodal data infrastructure, the implementation of closed-loop analytical systems, and the combination of experimental and attribution-based evaluation approaches. The necessity of aligning algorithmic optimization with long-term creative and brand strategies has been determined. Prospects for further research are associated with the development of methods for quantitative evaluation of multimodal content effectiveness, improvement of model interpretability, and investigation of the scalability of automated systems under conditions of dynamic digital environments.

Keywords: multimodal analysis, audiovisual content, data analytics, audience engagement, content optimization, digital marketing, behavioral metrics, adaptive systems.

Автоматизація тестування креативних гіпотез у відеомаркетингу на основі великих мовних моделей та комп'ютерного зору

Гладкова Марія Сергіївна,
магістр, старший відеограф і монтажер,
Golden Pet Brands, El Segundo, США
https://orcid.org/0009-0005-3966-2540

Анотація. Актуальність дослідження зумовлена зростанням ролі відеомаркетингу в цифровому середовищі, ускладненням структури аудіовізуального контенту та підвищенням вимог до обґрунтованості управлінських рішень, що обмежує ефективність традиційних підходів до тестування креативних рішень і актуалізує необхідність впровадження автоматизованих аналітичних систем. **Метою дослідження** визначено розроблення цілісної методичної основи автоматизації тестування креативних гіпотез у відеомаркетингу на основі великих мовних моделей та комп'ютерного зору з метою підвищення валідності, ефективності та масштабованості процесів прийняття рішень на основі даних. **Методи.** У процесі дослідження застосовано методи аналізу і синтезу для узагальнення теоретичних підходів до автоматизації тестування креативних гіпотез, систематизації для структурування їх компонентів і функціональних характеристик, порівняльного аналізу для оцінювання можливостей великих мовних моделей і комп'ютерного зору та логічного узагальнення для формування методичних підходів і практичних рекомендацій. **Результати.** Досліджено сутність креативних гіпотез як багатовимірних аналітичних конструкцій, що поєднують семантичні, візуальні та поведінкові параметри контенту. Виявлено, що використання мультимодальних технологій забезпечує перехід до безперервного тестування та адаптивної оптимізації відеоконтенту. Доведено, що інтеграція великих мовних моделей і комп'ютерного зору в єдину аналітичну систему дозволяє підвищити точність оцінювання ефективності контенту, скоротити витрати на експерименти та зменшити часові лаги між створенням і аналізом контенту. Встановлено, що ефективність автоматизації обмежується проблемами інтерпретованості моделей, якості та узгодженості даних, а також недосконалістю систем оцінювання результативності. **Висновки.** Обґрунтовано, що підвищення ефективності автоматизованого тестування досягається за рахунок формування інтегрованої мультимодальної інфраструктури даних,

 ISSN 3041-2129

впровадження замкнених аналітичних контурів та поєднання експериментальних і атрибутивних підходів до оцінювання. Визначено необхідність узгодження алгоритмічної оптимізації з довгостроковими креативними та брендовими стратегіями. Перспективи подальших досліджень пов’язані з розробленням методів кількісного оцінювання ефективності мультимодального контенту, підвищенням інтерпретованості моделей, а також дослідженням масштабованості автоматизованих систем у динамічних умовах цифрового середовища.

Ключові слова: мультимодальний аналіз, аудіовізуальний контент, аналітика даних, залученість аудиторії, оптимізація контенту, цифровий маркетинг, поведінкові метрики, адаптивні системи.

Problem statement. The rapid expansion of video marketing in digital environments has significantly intensified the need for systematic validation of creative hypotheses, as the effectiveness of visual narratives, editing patterns, and semantic messaging increasingly determines audience engagement and conversion outcomes – yet in practice, such validation remains largely fragmented, heuristic, and weakly formalized. Traditional A/B testing approaches are limited by high production costs, long feedback cycles, and the inability to isolate the contribution of specific creative elements within complex audiovisual content – this creates a methodological gap between the scale of available data and the capacity to generate statistically and economically grounded decisions. At the same time, the emergence of large language models and computer vision technologies enables automated interpretation of textual, visual, and multimodal signals, opening new possibilities for constructing adaptive systems of hypothesis generation, testing, and refinement in real time – however, their integration into coherent methodological frameworks for video marketing remains insufficiently developed. From a scientific perspective, the problem lies in the lack of unified approaches to modeling the relationship between creative features of video content and measurable performance indicators, as well as in the absence of validated metrics that would capture both semantic and perceptual dimensions of audience response – this constrains the development of reproducible and scalable analytical models. From a practical standpoint, businesses face increasing pressure to optimize content production under conditions of high market volatility and attention scarcity, which requires reducing experimentation costs while simultaneously increasing the precision and speed of decision-making – without automated tools, these objectives are difficult to achieve. Therefore, the development of integrated approaches to automating the testing of creative hypotheses in video marketing based on large language models and computer vision constitutes a relevant scientific and applied task, aimed at bridging the gap between data-driven analytics and creative strategy optimization.

Analysis of recent research and publications. A review of contemporary research demonstrates the formation of an interdisciplinary approach to the automation of creative hypothesis testing in video marketing, combining strategic digital marketing, large language models, and computer vision. V. Volikov and K. Shumilkina analyze the impact of strategic SEO on the digital visibility of international IT companies, which is essential for understanding how creative video hypotheses should align with broader marketing communication systems [1]. K. S. Mekhovych explores the transformation of the theory of innovative marketing effectiveness under conditions of uncertainty, emphasizing the need for flexible approaches to evaluating marketing decisions [2]. V. S. Rudenko and co-authors examine the integration of digital management and digital marketing as a factor in enhancing company competitiveness, thereby forming a managerial foundation for implementing automated systems for testing creative solutions [3]. O. V. Oliinyk and K. O. Shikovets substantiate the innovative component of enterprise development based on digital marketing management technologies, highlighting the role of digital tools in marketing decision-making processes [4].

The technological dimension of the problem is further elaborated in studies focusing on large language models, computer vision, and multimodal analysis. Q.

 ISSN 3041-2129

Yang and co-authors demonstrate the potential of combining predictive models with large language models to generate actionable recommendations in creative marketing, directly relating to the automated selection and validation of creative hypotheses [5]. M. Hasan and co-authors analyze applications of computer vision in the media and entertainment industry, revealing the potential of visual analytics for interpreting video content, scenes, objects, and visual patterns [6]. Z. Yang and co-authors investigate the capabilities of large language models for automated open-domain scientific hypothesis discovery, providing a methodological basis for transferring hypothesis generation and validation logic into video marketing contexts [7]. T. He and co-authors propose the DesignMinds approach, where video-based design ideation is enhanced through the integration of a vision-language model and a context-injected large language model, closely aligning with tasks of automated analysis of creative video solutions [8].

A distinct research stream focuses on the creative potential of artificial intelligence (AI) in marketing and content generation. D. Li and co-authors propose the Aug-Creativity framework for human-centered creativity using vision-language models, emphasizing the importance of balancing automation with human evaluation of creativity [9]. Z. Chen and J. Chan analyze the role of large language models in creative work, demonstrating that effectiveness depends on collaboration modality and user expertise [10]. P. Li and co-authors examine the validity of large language models for automated perceptual analysis, which is crucial for assessing audience responses to visual and semantic characteristics of video advertising [11]. L. Sun and co-authors compare the effectiveness of advertising copy generated by humans and Chinese large language models, highlighting both the capabilities and limitations of AI in marketing content generation [12].

At the same time, several studies address practical barriers and strategic implications of AI adoption in video and creative marketing. T. Yu and co-authors identify key barriers to the industry adoption of AI video generation tools from the perspective of video production professionals, emphasizing technological, professional, and organizational constraints [13]. M. Pagani and Y. Wind explore how artificial intelligence can unlock marketing creativity, positioning AI not only as a tool for idea generation but also as a mechanism for transforming the creative process itself [14]. C. X. Liang and co-authors provide a comprehensive overview of multimodal large language models in vision-language tasks, establishing a technical foundation for automated video content analysis through the integration of textual, visual, and contextual signals [15]. Collectively, these studies confirm that the automation of creative hypothesis testing in video marketing requires the integration of marketing analytics, multimodal machine intelligence, and expert interpretation to ensure both effectiveness and reliability of outcomes.

Identification of unresolved parts of the problem. Despite the rapid development of AI driven tools in digital marketing, a number of critical aspects of automating creative hypothesis testing in video marketing remain insufficiently resolved. In particular, the absence of a coherent methodological framework for linking semantic, visual, and behavioral parameters of content with measurable performance outcomes limits the interpretability and analytical reliability of results. Existing research is characterized by fragmentation and a focus on either technological capabilities or aggregated metrics, which prevents a comprehensive understanding of causal relationships and reduces the reproducibility of findings.

The importance of addressing these gaps lies in their direct impact on the validity and effectiveness of data-driven decision-making in dynamic digital environments. Their resolution enables the transition from descriptive analytics to systematic, scalable optimization of creative content. In this context, the study is aimed at refining conceptual and methodological foundations of automated testing and at establishing an integrated analytical approach that ensures consistent evaluation, reduces uncertainty, and enhances the adaptability of video marketing systems.

The purpose of the article is to develop an integrated methodological framework for automating the testing of creative hypotheses in video marketing

 ISSN 3041-2129

based on large language models and computer vision in order to enhance the validity, efficiency, and scalability of data-driven decision-making.

Objectives of the article:

1. To clarify the conceptual essence of creative hypotheses in video marketing and to analyze the capabilities of multimodal AI technologies in content evaluation.

2. To substantiate methodological approaches to integrating multimodal analytical tools into automated testing systems.

3. To identify key limitations of automation and to develop recommendations for improving system effectiveness.

Main material. The conceptual essence of creative hypotheses in video marketing lies in their function as formally articulated assumptions regarding the relationship between specific elements of audiovisual content and measurable audience responses – including engagement, retention, and conversion. Structurally, such hypotheses integrate semantic, visual, temporal, and behavioral components, forming a multidimensional construct that reflects both the logic of content production and the mechanisms of user perception. Unlike intuitive creative decisions, hypothesis-driven approaches enable the decomposition of video content into analytically tractable elements, which can be systematically tested and optimized through data-driven methods. This transformation is particularly relevant under conditions of high content saturation and algorithmically mediated distribution, where even minor variations in narrative structure, visual composition, or message framing can significantly affect performance outcomes (table 1).

Table 1

Structural characteristics of creative hypotheses in video marketing

Component	Substantive characteristic	Key elements of manifestation	Functional role in effectiveness formation
Semantic	Definition of core message and value proposition	Script logic, key message framing, call-to-action formulation	Ensures clarity and relevance of communication

 ISSN 3041-2129

Component	Substantive characteristic	Key elements of manifestation	Functional role in effectiveness formation
Visual	Representation of content through imagery and design	Color schemes, composition, object focus, branding elements	Influences attention capture and emotional response
Temporal	Organization of content in time	Video length, pacing, scene transitions, timing of key moments	Determines retention dynamics and viewing completion
Behavioral	Anticipation of audience reaction patterns	Click behavior, interaction triggers, engagement signals	Links content features to measurable user actions
Contextual	Adaptation to platform and audience specifics	Platform algorithms, audience segmentation, content format requirements	Aligns hypothesis with distribution environment constraints

Source: developed by the author based on [1; 2; 3; 14; 10, p. 9105; 12, p. 12; 13].

In practical application, the presented structural decomposition allows creative hypotheses to be transformed into measurable and testable design parameters embedded directly into the production and distribution cycle of video content. For example, the semantic component is operationalized through parallel script variations, where alternative value propositions or call-to-action formulations are systematically encoded into multiple video versions – in performance campaigns this often manifests as testing distinct narrative hooks within the first 3–5 seconds to determine which semantic entry point maximizes audience retention. The visual component is increasingly quantified using computer vision tools that assess object saliency, color contrast, and frame density – in practice, brands optimize foreground focus or visual clutter based on detected correlations with drop-off rates, particularly in mobile-first environments where attention windows are limited [10, p. 9105]. The temporal component is directly linked to micro-level editing decisions – such as cut frequency or rhythm alignment with platform-specific consumption patterns – where even slight adjustments in pacing can alter completion rates; for instance, shortening transition intervals in short-form videos often leads to measurable increases in watch-through metrics. The behavioral component is realized through the integration of event-level analytics, where user actions (clicks, pauses, replays) are mapped back to specific content segments, enabling the identification of causal relationships rather than aggregate correlations. Finally, the contextual component ensures that hypotheses are not evaluated in isolation but are adapted to algorithmic distribution logics – for example, content variants optimized for one platform's recommendation system may underperform in another due to differences in ranking signals and audience interaction models [13].

Such an approach enables a shift from post hoc evaluation toward embedded experimentation, where creative decisions are continuously recalibrated through feedback loops combining large language model–generated variations and computer vision–based feature extraction. In contemporary marketing practice, this results in the formation of iterative content pipelines, where hypotheses are not only tested but dynamically refined across successive deployment cycles, reducing the cost of error and increasing the precision of targeting under conditions of high market volatility.

The capabilities of large language models and computer vision in multimodal video content are grounded in their capacity to jointly process semantic and visual information streams, enabling the generation, structured interpretation, and predictive evaluation of content effectiveness within a unified analytical framework (table 2).

Table 2

Capabilities of large language models and computer vision in multimodal video content processing

Functional domain	Technological manifestation	Analytical outputs	Applied value in video marketing
Content synthesis	Generation of scripts and narrative structures by language models	Textual variants, semantic clusters, tone differentiation	Enables rapid creation of diverse content scenarios aligned with target segments

ISSN 3041-2129

Functional domain	Technological manifestation	Analytical outputs	Applied value in video marketing
Visual decoding	Automated recognition of objects, scenes, and motion patterns	Feature maps, object hierarchies, scene segmentation	Provides structured representation of visual composition for further analysis
Cross-modal alignment	Integration of textual and visual representations into unified embeddings	Multimodal vectors, semantic consistency scores	Ensures coherence between message and visual delivery
Predictive evaluation	Modeling expected audience response based on historical and real-time data	Engagement forecasts, retention probabilities, conversion likelihood	Supports proactive optimization of content before large-scale release
Adaptive refinement	Iterative adjustment of content based on feedback signals	Parameter updates, variant ranking, performance gradients	Enables continuous improvement of content through automated feedback loops

Source: developed by the author based on [5, p. 10; 6, p. 210; 7, p. 13550; 8, p. 6; 11, p. 258; 15].

In applied settings, these functions are implemented as a continuous analytical-production loop rather than isolated tools, where each stage directly informs subsequent decisions. For instance, content synthesis is no longer limited to drafting scripts but involves generating controlled semantic variations tied to specific hypotheses about audience motivation – such as contrasting rational versus emotional appeals within identical product narratives. These variants are immediately passed to visual decoding systems, which quantify scene composition parameters, including object prominence, background complexity, and motion intensity, allowing the identification of visual patterns that correlate with attention retention in comparable datasets [7, p. 13550]. Cross-modal alignment becomes critical at this stage, as inconsistencies between textual intent and visual execution frequently lead to performance degradation – in practice, campaigns often reveal that high-performing scripts lose effectiveness when paired with visually overloaded

 ISSN 3041-2129

or semantically ambiguous frames. Multimodal embeddings enable the detection of such mismatches before full deployment by measuring coherence scores between narrative tone and visual structure. Predictive evaluation further extends this logic by integrating historical campaign data with real-time signals, allowing the estimation of engagement trajectories for each variant – for example, predicting early drop-off risks in videos where key informational cues appear too late relative to typical user interaction patterns [15]. Adaptive refinement operates during live campaign execution, where interaction-level data are continuously mapped to specific content features, enabling targeted adjustments rather than full content replacement. In practice, this may involve selectively modifying opening frames, altering pacing in underperforming segments, or dynamically prioritizing variants with higher predicted conversion potential across different audience clusters [8, p. 6]. As a result, multimodal technologies support a transition toward granular, evidence-based content optimization, where creative decisions are iteratively validated and refined within a tightly integrated analytical environment, reducing uncertainty while preserving the flexibility required in rapidly changing digital markets.

The methodological integration of multimodal analytical tools into automated systems for testing creative hypotheses in video marketing is based on the coordination of heterogeneous data, model interactions, and evaluation logic within a unified experimental cycle that ensures analytical consistency and operational scalability (table 3).

Table 3

Methodological approaches to integrating multimodal analytical tools into automated systems for testing creative hypotheses in video marketing

Integration layer	Methodological principle	Operational mechanism	Resulting analytical effect
Data harmonization	Standardization of heterogeneous	Unified data schemas, temporal alignment of text–	Eliminates inconsistencies and enables cross-modal comparability

Integration layer	Methodological principle	Operational mechanism	Resulting analytical effect
	inputs across modalities	video–interaction streams	
Feature orchestration	Coordinated extraction and structuring of multimodal attributes	Joint feature spaces, synchronized embeddings, dimensionality reduction	Forms a coherent analytical basis for hypothesis testing
Model coupling	Integration of multiple analytical models within a single pipeline	Sequential and parallel model architectures, inference chaining	Enhances depth and robustness of analytical conclusions
Evaluation design	Formalization of metrics and testing logic	Multi-level KPIs, experimental control groups, causal inference methods	Ensures validity and interpretability of results
Feedback governance	Regulation of iterative updates based on performance signals	Adaptive thresholds, reinforcement loops, automated re-weighting of variables	Supports continuous system learning and optimization

Source: developed by the author based on [3; 5, p. 15; 7, p. 13555; 8, p. 10; 11, p. 260; 13; 15].

The effectiveness of these approaches is determined by how precisely the system links granular content attributes with observable behavioral outcomes across the full lifecycle of video deployment. Data harmonization, for instance, is not limited to technical alignment but enables frame-level attribution, where specific narrative elements or visual transitions can be directly associated with retention dips or interaction spikes – in practice, this allows marketers to isolate micro-level inefficiencies such as delayed message exposure or visually overloaded segments [11, p. 260]. Feature orchestration extends this by constructing analytically meaningful representations, where combined indicators such as semantic intensity and visual contrast reveal nonlinear effects, for example cases in which moderate visual simplification amplifies the impact of a strong narrative hook rather than diminishing engagement. Model coupling introduces operational continuity between generation and evaluation, eliminating the traditional gap between content creation and analytics – a typical implementation involves automatic routing of generated scripts into predictive modules that estimate performance under different audience segments, enabling early-stage filtering of low-potential variants before resource-intensive production [7, p. 13555]. Evaluation design ensures that such filtering remains empirically grounded – organizations embed controlled exposure mechanisms into distribution platforms, where comparable audience cohorts are used to test alternative variants, allowing causal interpretation of performance differences rather than reliance on aggregate metrics distorted by algorithmic bias. Feedback governance operationalizes learning by translating performance signals into parameter adjustments at the system level – instead of manual iteration, the system dynamically recalibrates weighting schemes for features such as pacing, visual density, or semantic framing based on their marginal contribution to target metrics. In contemporary practice, this enables the formation of adaptive experimentation environments in which creative hypotheses are continuously refined across successive deployment cycles, reducing uncertainty while maintaining the flexibility required to respond to rapid shifts in audience behavior and platform algorithms.

The automation of creative hypothesis testing in video marketing exposes a set of tightly interrelated scientific and practical problems arising from the complexity of multimodal data and algorithmic inference. A key issue is limited interpretability, as large language models and computer vision systems generate high-dimensional representations that obscure causal relationships between specific creative elements and performance outcomes – this constrains the transparency, reproducibility, and theoretical generalization of results [11, p. 258].

Data reliability constitutes another critical challenge, since multimodal inputs are heterogeneous, temporally misaligned, and often incomplete. Text, visual content, and interaction signals are produced under different conditions, leading to structural inconsistencies and biased representations of user behavior – in dynamic digital environments, this problem is intensified by continuous changes in platform algorithms and audience patterns, reducing the stability of empirical regularities.

Significant limitations are also observed in evaluation metrics, which remain largely aggregated and insufficiently sensitive to the multidimensional nature of creative effectiveness. Indicators such as engagement or retention do not isolate the contribution of individual content features, creating ambiguity in hypothesis validation and limiting cross-platform comparability. At the same time, the absence of standardized multimodal benchmarks complicates cumulative knowledge development.

From a modeling perspective, difficulties arise in cross-modal integration and feature attribution, where combining semantic and visual representations leads to information loss and increased risk of overfitting, particularly under limited or non-representative data conditions [6, p. 210]. Automated feedback loops may further reinforce historical biases, narrowing the diversity of creative solutions and reducing adaptive capacity over time.

Practical implementation is constrained by high computational demands, latency in real-time processing, and the need to integrate analytical systems into existing organizational workflows. Additional challenges relate to data governance, including privacy, consent, and algorithmic accountability, as well as to the strategic risk of over-optimization, where short-term performance gains may conflict with long-term brand coherence and creative differentiation.

Improving the effectiveness of automated systems for testing creative hypotheses in video marketing under dynamic digital conditions requires the alignment of analytical precision with operational flexibility and controlled experimentation logic. At the system level, priority should be given to the implementation of modular architectures in which data processing, model inference, and decision execution are functionally decoupled yet synchronised through unified protocols – this enables rapid reconfiguration of analytical pipelines without disrupting ongoing campaigns. Equally important is the establishment of

ISSN 3041-2129

standardized multimodal data frameworks that ensure temporal alignment and semantic consistency across textual, visual, and interaction streams, thereby reducing structural noise and increasing the reliability of analytical outputs.

From a methodological perspective, it is advisable to incorporate hybrid evaluation models that combine aggregate performance indicators with feature-level attribution metrics, allowing the identification of causal relationships between specific content elements and user responses. The integration of controlled experimental designs directly into distribution environments should be treated as a baseline requirement, ensuring that automated decisions are grounded in comparable exposure conditions rather than platform-driven variability. At the same time, adaptive threshold mechanisms should be introduced to regulate model sensitivity to performance fluctuations, preventing premature optimization based on statistically unstable signals.

In practical implementation, the effectiveness of such systems is enhanced through the use of iterative content pipelines, where generated variations are pre-filtered using predictive models before full deployment, reducing production costs and experimental overload. Continuous feedback loops must be complemented by governance mechanisms that limit bias amplification, for example through periodic re-initialization of model parameters or diversification constraints in content generation. Particular attention should be paid to latency reduction by optimizing data flow architectures and prioritizing near-real-time analytics for high-impact decision points, such as early-stage audience retention.

Organizationally, the integration of automated testing systems requires the formalization of interaction between creative and analytical functions, where hypotheses are explicitly encoded, monitored, and iteratively refined within shared operational frameworks. Finally, ensuring long-term effectiveness necessitates balancing algorithmic optimization with strategic brand consistency, which can be achieved by embedding qualitative control criteria into automated decision rules –

this prevents the erosion of narrative coherence while maintaining responsiveness to rapidly changing audience behavior and platform dynamics.

Conclusions. It has been established that the automation of creative hypothesis testing in video marketing based on large language models and computer vision enables the transition from intuitive content design to structured, data-driven experimentation, where audiovisual elements are systematically decomposed and evaluated within continuous feedback cycles. This ensures higher precision in identifying performance drivers and supports predictive and adaptive optimization of content in dynamic digital environments. The study identifies key limiting factors, including low interpretability of multimodal models, heterogeneity and instability of data, and insufficient sensitivity of evaluation metrics to complex creative effects. Additional constraints arise from fragmented methodological approaches, risks of bias amplification, and high implementation complexity, which collectively reduce the reliability and scalability of automated systems. It has been substantiated that improving effectiveness requires integrated data infrastructures, closed-loop analytical architectures, and the combination of experimental design with feature-level attribution. Priority should be given to adaptive governance mechanisms and the alignment of algorithmic optimization with strategic brand coherence. Further research should focus on developing quantitative evaluation models for multimodal content, enhancing interpretability and causal analysis, and ensuring scalability of automated systems under conditions of rapidly evolving digital platforms and audience behavior.

References

1. Воліков В., Шумілкіна К. Вплив стратегічного SEO на цифрову видимість міжнародних IT-компаній у системі маркетингових комунікацій. *Вісник Хмельницького національного університету. Серія: Економічні науки*. 2026. Вип. 350, № 1. С. 579–588. DOI: https://doi.org/10.31891/2307-5740-2026-350-79.

2. Мехович К. С. Трансформація теорії ефективності інноваційного маркетингу в умовах невизначеного майбутнього. *Енергозбереження. Енергетика. Енергоаудит*. 2025. Вип. 1, № 204. С. 79–95. DOI: https://doi.org/10.20998/2313-8890.2025.01.07.

3. Руденко В. С., Марухленко О. В., Лисенко І. В. Інтеграція digital management і цифрового маркетингу у підвищенні конкурентоспроможності компаній. *Економіка та суспільство*. 2025. № 78. DOI: https://doi.org/10.32782/2524-0072/2025-78-156.

4. Олійник О. В., Шіковець К. О. Інноваційна складова розвитку підприємств на основі цифрових технологій маркетингового управління. *Актуальні проблеми сталого розвитку*. 2025. Вип. 2, № 5. С. 130–137. DOI: https://doi.org/10.60022/2(5)-16S.

5. Yang Q., Farseev A., Ongpin M. et al. Fusing predictive and large language models for actionable recommendations in creative marketing. *ACM Transactions on Information Systems*. 2025. Vol. 43, № 5. P. 1–31. DOI: https://doi.org/10.1145/3725885.

6. Hasan M., Athrey K. S., Khalid A. et al. Applications of computer vision in entertainment and media industry. In: *Computer Vision*. Chapman and Hall/CRC. 2024. P. 205–238. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003328957-10/applications-computer-vision-entertainment-media-industry-mahmudul-hasan-kishan-shamsundar-athrey-arfeen-khalid-danfeng-xie-ehsan-younessian-tony-braskich (accessed: 04.01.2026).

7. Yang Z., Du X., Li J. et al. Large language models for automated open-domain scientific hypotheses discovery. In: *Findings of the Association for Computational Linguistics: ACL 2024*. 2024. P. 13545–13565. DOI: https://doi.org/10.18653/v1/2024.findings-acl.804.

8. He T., Stanković A., Niforatos E., Kortuem G. DesignMinds: Enhancing Video-Based Design Ideation with a Vision-Language Model and a Context-Injected

Large Language Model. In: *Proceedings of the 7th ACM Conference on Conversational User Interfaces*. 2025. P. 1–15. DOI: https://doi.org/10.1145/3719160.3736633.

9. Li D., Xia L., Fan L. Aug-Creativity: Framework for Human-Centered Creativity with Vision Language Models. In: *International Conference on Human-Computer Interaction*. Cham: Springer Nature Switzerland. 2025. P. 86–101. DOI: https://doi.org/10.1007/978-3-031-93864-1_7.

10. Chen Z., Chan J. Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*. 2024. Vol. 70, № 12. P. 9101–9117. DOI: https://doi.org/10.1287/mnsc.2023.03014.

11. Li P., Castelo N., Katona Z., Sarvary M. Determining the validity of large language models for automated perceptual analysis. *Marketing Science*. 2024. Vol. 43, № 2. P. 254–266. DOI: https://doi.org/10.1287/mksc.2023.0454.

12. Sun L., An Y., Song X., Liu J. Human vs. AI in creative contests: a study on the effectiveness of advertisement copy generated by Chinese large language models. *Electronic Commerce Research*. 2025. P. 1–43. DOI: https://doi.org/10.1007/s10660-025-09990-2.

13. Yu T., Yang W., Xu J., Pan Y. Barriers to industry adoption of AI video generation tools: A study based on the perspectives of video production professionals in China. *Applied Sciences*. 2024. Vol. 14, № 13. Article 5770. DOI: https://doi.org/10.3390/app14135770.

14. Pagani M., Wind Y. Unlocking marketing creativity using artificial intelligence. *Journal of Interactive Marketing*. 2025. Vol. 60, № 1. P. 1–24. DOI: https://doi.org/10.1177/10949968241265855.

15. Liang C. X., Tian P., Yin C. H. et al. A comprehensive survey and guide to multimodal large language models in vision-language tasks. 2024. *arXiv preprint*. arXiv.2411.06284. DOI: https://doi.org/10.48550/arXiv.2411.06284.