Científica

vol. 29, no. 2, July-December. 2025, pp. 01-11
ISSN 2594-2921, Instituto Politécnico Nacional, MÉXICO
e290208 | DOI: 10.46842/ipn.cien.v29n2a08

ARTICLE

A Review of Multimodal Artificial Intelligence: Ethical Challenges and Practical Benefits in Content Generation

Una revisión sobre Inteligencia artificial multimodal: Retos éticos y beneficios prácticos en la generación de contenidos

https://cientifica.site

German **Cuaya Simbro** [1]
Wendy **Garcia Tapia** [2]
María F. **Avila Vargas** [3]
Pedro Aarón **Hernández Ávalos** [4]

Tecnológico Nacional de México
ITS del Oriente del Estado de Hidalgo, Hidalgo, MÉXICO
[1] ORCID: 0000-0001-6303-154X / gcuaya@itesa.edu.mx
[2] ORCID: 0009-0009-9401-3892 / 24030631m@itesa.edu.mx
[3] ORCID: 0009-0005-7612-7897 / 24030632m@itesa.edu.mx

OMB Services S de RL de CV, Puebla, MÉXICO
[4] ORCID: 0009-0006-2880-191X / pedro.aaron@gmail.com

Recibido 06/08/2025, aceptado 08/12/2025.

Abstract

This systematic review examines 23 peer-reviewed studies published since 2020 to delineate technical and ethical risks, synthesize key benefits, and identify future research avenues for AI-driven multimodal content creation. Employing a PRISMA-guided protocol, we screened five major digital libraries using rigorous inclusion/exclusion criteria and an adapted CASP checklist. The resulting corpus encompasses diverse domains, including visual art, education, marketing, healthcare, and disaster response. We classified these works into four functional categories: Creation and Design, Communication and Analysis, Automation and Detection, and Interaction and Teaching. Quantitative analysis reveals that 30% of studies prioritized automation efficiency, 26% highlighted personalized outputs, and 18% reported enhanced content diversity. Crucially, our review identifies significant risks such as misinformation, modality-alignment failures, algorithmic bias, and privacy breaches. These findings underscore the need for transparent algorithms, bias-monitoring protocols, and privacy-by-design frameworks. We conclude by advocating for the development of interpretable models and standardized ethical methodologies to advance robust, responsible, and scalable multimodal AI systems.

Index terms: multimodal AI, AI-generated content, ethical challenges, automation efficiency.

Resumen

Esta revisión sistemática examina 23 estudios revisados por pares publicados desde 2020 para delinear los riesgos técnicos y éticos, sintetizar los beneficios clave e identificar futuras vías de investigación para la creación de contenido multimodal impulsada por IA. Empleando un protocolo guiado por PRISMA, examinamos cinco bibliotecas digitales principales utilizando rigurosos criterios de inclusión y exclusión, junto con una lista de verificación CASP adaptada. El corpus resultante abarca diversos dominios, incluyendo artes visuales, educación, marketing, salud y respuesta ante desastres. Clasificamos estos trabajos en cuatro categorías funcionales: Creación y Diseño, Comunicación y Análisis, Automatización y Detección, e Interacción y Enseñanza. El análisis cuantitativo revela que el 30% de los estudios priorizaron la eficiencia de la automatización, el 26% destacaron los resultados personalizados y el 18% reportaron una mayor diversidad de contenido. Crucialmente, nuestra revisión identifica riesgos significativos como la desinformación, fallos en la alineación de modalidades, sesgo algorítmico y violaciones de privacidad. Estos hallazgos subrayan la necesidad de algoritmos transparentes, protocolos de monitoreo de sesgos y marcos de privacidad desde el diseño. Concluimos abogando por el desarrollo de modelos interpretables y metodologías éticas estandarizadas para avanzar hacia sistemas de IA multimodal robustos, responsables y escalables.

Palabras clave: IA multimodal, contenido generado por IA, retos éticos, automatización eficiente.

I. Introduction

Artificial Intelligence (AI) has rapidly evolved into a comprehensive platform capable of generating and integrating heterogeneous media—text, visual imagery, audio signals, video streams, and augmented-reality overlays — thereby enabling unprecedented versatility in content creation. Multimodal AI, which integrates multiple data types to enhance processing, understanding, and content generation, has attracted considerable attention due to its potential to advance applications in design, production assistance, accessibility tools, and education [1]. Compared to unimodal frameworks, multimodal architectures yield markedly enhanced contextual awareness and facilitate deeper user–system interactions, features that are indispensable for addressing complex tasks in domains such as human-computer interaction and assistive technologies [1], [2].

The evolution of multimodal AI can be traced back to early attempts at integrating text and image processing in the 1990s. Over the past decade, advancements in deep learning and neural networks have significantly enhanced the capabilities of multimodal systems, enabling more sophisticated and innovative applications across various domains, design, accessibility, and education.

However, this rapid expansion in capabilities has not been devoid of obstacles. Although multimodal artificial intelligence has shown enormous potential, it currently faces significant challenges, such as the integration and interpretation of multiple types of data. These challenges stem from the unique characteristics of each modality and the substantial computational resources required for efficient processing. In the ethical field, the creation of multimodal AI raises concerns about bias in training data, and authenticity, the above underlines the importance of conducting research that explores solutions to achieve a safe, efficient, and ethical integration.

To address these complexities, this study presents a systematic review of the literature published between 2020 and 2024. The primary objective is to (1) delineate the principal technical and ethical risks, (2) synthesize key benefits observed across application domains, and (3) identify prospective research avenues. Unlike previous surveys that isolate specific modalities, this review adopts a holistic perspective on multimodal content creation, analyzing how the integration of text, image, and audio is reshaping industries.

Following the PRISMA methodology, we analyzed 23 peer-reviewed studies. Our preliminary analysis indicates a paradigm shift: while early research focused on multimodal classification and detection, recent literature is heavily dominated by Generative AI (GenAI) and its implications for creative labor. This review contributes a categorized synthesis of these findings, offering a dual perspective on the efficiency gains and the ethical debt accrued by these rapid advancements.

II. Methodology

We conducted a systematic review following the PRISMA [3] methodology to identify and synthesize key findings and empirical results from peer-reviewed studies on AI-driven multimodal content creation. Figure 1 shows the steps followed, which are explained in detail in the next sections.

Científica, vol. 29, no. 2, pp. 01-11, July-December 2025,
ISSN 2594-2921, Instituto Politécnico Nacional, México
e290208 | DOI: 10.46842/ipn.cien.v29n2a08

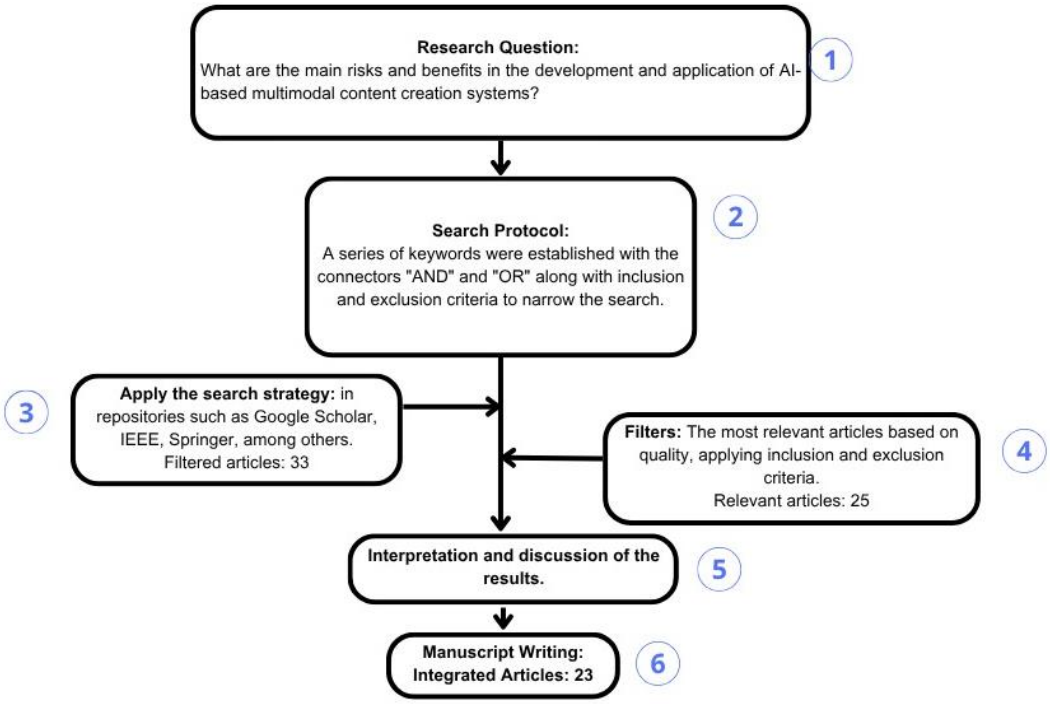

Fig. 1. Flowchart of the steps of the systematic review.

A. Determining the Research Question

The research question was established to delimit the context to be investigated, resulting in guiding the research, it was: What are the potential risks and benefits of AI-based multimodal content creation systems, and how can they be balanced to ensure ethical and effective use?

B. Develop a search protocol

We defined the search protocol using the following Boolean expression: "("multimodal artificial intelligence" OR "AI multimodal content creation") AND ("AI-generated content" OR "digital content with multimodal AI") AND ("opportunities" OR "challenges" OR "risks") AND "multimodal deep learning".

C. Apply the search strategy

A comprehensive search strategy was implemented using databases including IEEE Xplore, SpringerLink, Elsevier ScienceDirect, Scopus, and Web of Science, resulting in an initial retrieval of 33 articles. These were evaluated using predefined inclusion and exclusion criteria designed to ensure methodological rigor and thematic relevance.

- **Inclusion criteria were:** i) peer-reviewed studies and systematic reviews published from January 2020 onward; ii) empirical case studies evaluating AI-based multimodal content creation and its effects on user engagement; and iii) explicit discussion of technical or ethical risks.

– **Exclusion criteria** ruled out studies published prior to 2020, as well as theses, conference papers, and articles on content creation that did not involve a multimodal AI focus.

We applied an adapted Critical Appraisal Skills Programme (CASP) checklist [4] to assess methodological transparency, research relevance, and result validity. 8 articles were excluded for lacking a multimodal AI focus, and 2 were discarded for failing to meet our quality threshold, yielding 23 studies for final synthesis.
Key information from each article, including title, objectives, and significant findings, was systematically extracted and compiled in a comparative table. All selected studies demonstrated the application of multimodal AI in various domains and provided valuable insights into the benefits and potential risks associated with its use in content creation, particularly in the context of social media interaction.

D. Interpretation and discussion of the results

We analyzed our findings in the context of existing multimodal AI literature, identifying key application domains for AI-driven multimodal content creation.

Figure 2 presents a synthesized map of the 4 principal application domains for AI-driven multimodal content creation identified across the 23 reviewed studies. The extended discussion of the findings are discussed in section IV.

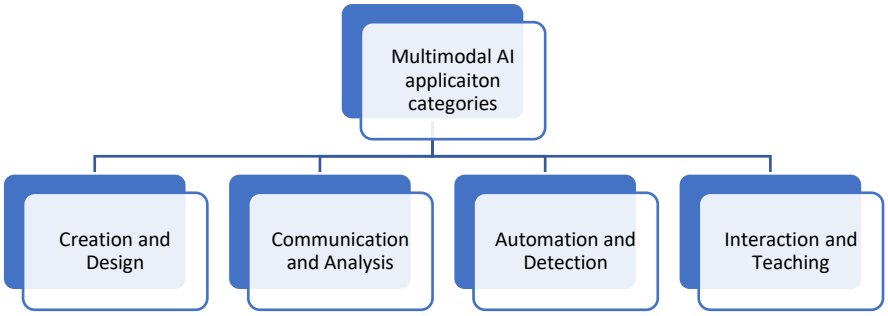

Fig. 2. Multimodal AI application category map.

Temporal Evolution of Multimodal Approaches Analysis of the selected corpus reveals a distinct temporal evolution in the application of multimodal AI. Studies published between 2020 and 2021 [22], [23], [25] predominantly focused on Automation and Detection tasks—such as tumor characterization or disaster response analysis—leveraging multimodal deep learning for classification accuracy.

Conversely, a significant shift occurred from 2023 onwards [5], [6], [11], [12], coinciding with the proliferation of diffusion models and Large Language Models (LLMs). The literature from 2023–2024 is heavily concentrated in the Creation and Design and Interaction and Teaching categories. This trajectory suggests that the field has moved from interpreting multimodal data to generating it, raising new ethical challenges regarding authorship and authenticity that were less prevalent in the earlier diagnostic-focused studies.

III. Results

The following section presents our key findings and critical discussion, synthesizing quantitative and qualitative results to address the objectives of this systematic review.

Table 1 presents a qualitative overview of ten key studies in AI-driven multimodal content creation, including each work's reference, title, and a concise summary of its study's primary findings. These works served as the foundation for our classification of the 4 principal application domains of multimodal models.

Table 1
Principal titles on artificial intelligence content creation and relevant results.

Research´s Title	Key Findings
Category: Creation and Design	
Generative AI tools: visual exploration with The House of Bernarda Alba, [6].	Didactic proposal for the creation of images and simulations, empowering students in their interaction with academic knowledge and their creative expression.
Visual artistic creation facing the challenges of artificial intelligence: creative automation and ethical questions, [8].	The risks of a devaluation of human creativity and a derivative visual culture.
Category: Communication and Analysis	
Impact and opportunities of artificial intelligence in visual art: personalization of creativity with stable diffusion and control net, [5].	New types of inputs both modal and multimodal, can improve the interaction between artists and generative tolos.
The impact of Generative Artificial Intelligence on the discipline of communication, [7].	Objective: to bring together emerging works from researchers who discuss how AI is transforming the creation of digital content.
Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI, [11].	Overview of state-of-the-art multimodal techniques.
Artificial intelligence and marketing: dangers and opportunities, [13].	Challenges and opportunities of AI in marketing, from creation to knowledge transfer.
Artificial Intelligence (AI): Multidisciplinary Perspectives on emerging challenges, and opportunities, [14].	Proposes a research agenda to address the effects of AI.
Category: Automation and Detection	
Multimodal AIBased Summarization and Storytelling for Soccer on social media, [10].	Generating football match highlights and summaries for social media.
Category: Interaction and Teaching	
Artificial Intelligence in Higher Education: opportunities and risks, [9].	Benefits that AI can bring to the educational area of Universities.
Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives, [12].	Personalized AI agent to interact with loved ones after death.

The 10 studies in Table 1 form the empirical backbone of our systematic review by exemplifying four core roles of AI-driven multimodal content creation. Creation and Design works (Studies 2 and 4) introduce novel input architectures and artist–AI collaboration techniques that shaped our criteria for evaluating user-centered generation methods, other research related [5], [15], [16]. Communication and Analysis contributions (Studies 1, 3, 7, 9, 10) offer comprehensive mappings of application areas, systematic surveys of prompting strategies, and sector-specific risk assessments that informed our classification schema and highlighted methodological gaps, within this category, we also include related research [17], [18], [19], [20] . Automation and Detection (Study 6) demonstrates practical

Científica, vol. 29, no. 2, pp. 01-11, July-December 2025,
ISSN 2594-2921, Instituto Politécnico Nacional, México
e290208 | DOI: 10.46842/ipn.cien.v29n2a08

pipelines for automated summarization and highlight extraction, providing concrete performance benchmarks for our efficiency-gain analysis (Table 1, Study 6), research related [21], [22], [23], [24], [25]. Finally, Interaction and Teaching studies (Studies 5 and 8) showcase adaptive learning environments and conversational AI agents, grounding our discussion of pedagogical benefits and interface design considerations, supported by other studies [26], [27], [28]. Together, these categorized works ensure our review captures the field's design innovations, analytical insights, automation capabilities, and interactive pedagogies.

From the studies reviewed, we identified and classified those that reported some type of benefit. Figure 3 highlights the key benefits identified in our sample. Regarding the frequency of benefits, analysis of the reviewed corpus indicates that 30% of the selected studies prioritized automation efficiency as the foremost advantage, followed by enhanced personalization (26%) and greater content diversity (18%). These findings demonstrate that multimodal architectures not only streamline the content-creation process but also deliver tailored outputs of consistently high quality [20].

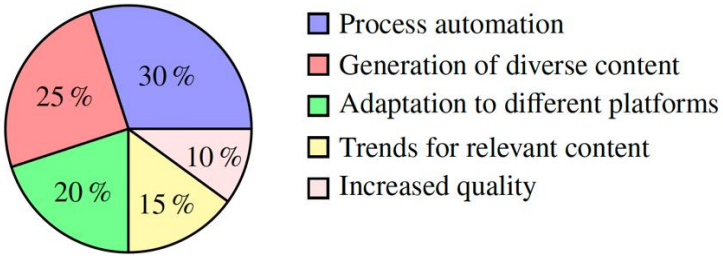

Fig. 3. Benefits of multimodal AI.

We identified several substantial risks associated with multimodal AI in content creation. Integrating heterogeneous data sources often leads to alignment and consistency challenges across modalities, which can degrade result quality [29]. Moreover, ethical concerns such as algorithmic bias and data privacy arise when models require large volumes of sensitive information [30]. As illustrated in Figure 4, the most frequently reported risks include misinformation and false content, modality-alignment failures, bias amplification, and privacy breaches. These findings underscore the necessity of developing robust mitigation approaches—namely, transparent algorithmic practices, continuous bias monitoring protocols, and stringent privacy-by-design policies during data collection and processing—to safeguard against the negative consequences of AI-generated multimodal content.

We identified several substantial risks associated with multimodal AI in content creation. Integrating heterogeneous data sources often leads to alignment and consistency challenges across modalities, which can degrade result quality [29]. Moreover, ethical concerns such as algorithmic bias and data privacy arise when models require large volumes of sensitive information [30]. As illustrated in Figure 4, the most frequently reported risks include misinformation and false content, modality-alignment failures, bias amplification, and privacy breaches. These findings underscore the necessity of developing robust mitigation approaches—namely, transparent algorithmic practices, continuous bias monitoring protocols, and stringent privacy-by-design policies during data collection and processing—to safeguard against the negative consequences of AI-generated multimodal content.

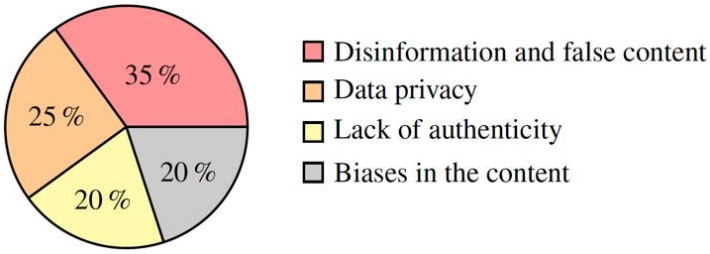

Fig. 4. Major risks of multimodal AI.

Our systematic review's categorization, Creation and Design, Communication and Analysis, Automation and Detection, and Interaction and Teaching, captures the multifaceted impact of multimodal AI on content creation. Creation and Design studies demonstrate how novel architectures empower richer, user-centered generation, while Communication and Analysis works reveal methodological gaps and inform robust classification frameworks. Automation and Detection research underscores efficiency gains through streamlined summarization and highlight extraction, and Interaction and Teaching investigations highlight the pedagogical value of adaptive learning environments and conversational agents. However, these benefits coexist with significant risks: alignment inconsistencies, misinformation, algorithmic bias, and privacy breaches. By systematically mapping these categories, advantages, and challenges, our review not only clarifies the current state of multimodal AI research but also identifies critical avenues for future studies aimed at enhancing methodological rigor, ethical governance, and practical deployment.

A. Methodological Limitations

Although this review followed a rigorous protocol based on PRISMA, some limitations must be acknowledged.

- The search was restricted to five major academic databases; while comprehensive, this may exclude relevant grey literature or preprints not yet indexed, which is significant in the fast-evolving field of GenAI.
- The inclusion criteria focused on articles published in English and Spanish, potentially overlooking significant contributions in other languages.
- The inclusion of the specific term "multimodal deep learning" in our search string may have introduced a bias towards deep neural network architectures, potentially excluding multimodal systems based on other computational frameworks. Future reviews should consider broader terminologies to capture the full spectrum of multimodal AI.
- Given the exponential pace of multimodal AI development, some technical constraints mentioned in studies from 2020 or 2024 may have already been superseded by newer models released during the review process.

IV. Discussion

This review synthesized the state of multimodal AI content creation, revealing a polarized landscape between technical optimization and ethical caution.

- Tension Between Efficiency and Ethics - A cross-analysis of the selected studies highlights a divergence in priorities. Papers focused on Automation and Communication [10], [13] tend to frame multimodal AI as a tool for "efficiency" and "personalization," often underreporting potential externalities. In contrast, studies in the Creation and Design category [8], [11] adopt a more critical stance, warning of a "derivative visual culture" and the devaluation of human agency. This contrast suggests that while the engineering perspective

prioritizes metrics (speed, accuracy), the humanities-oriented perspective prioritizes semantic integrity and cultural impact.

– Concepts and Future Trends - The progression from detection-based systems (2020) to generative agents (2024) implies that future research will no longer struggle with how to fuse modalities, but how to control the output. The concept of "Hallucination" in multimodal systems is emerging as a critical barrier. While unimodal text models hallucinate facts, multimodal models suffer from "grounding failures" where the image does not match the text prompt [11]. Our analysis suggests that the next wave of research must focus on Explainable Multimodal AI (XAI) to bridge the gap between the high-performance "black boxes" of deep learning and the user's need for trust and verification.

V. Conclusions

Our systematic review demonstrates that AI-driven multimodal content creation delivers significant advantages across four foundational roles, Creation and Design, Communication and Analysis, Automation and Detection, and Interaction and Teaching.

Simultaneously, our analysis identified critical risks that must be addressed to ensure responsible deployment: modality-alignment inconsistencies, misinformation and false content, algorithmic bias, and data-privacy breaches. These findings underscore the necessity of transparent algorithmic practices, continuous bias-monitoring protocols, and privacy-by-design frameworks throughout data acquisition and processing.

Looking forward, we recommend prioritizing: Interpretable and robust model architectures that facilitate auditability and user trust; Standardized, ethical data-handling methodologies that balance innovation with privacy safeguards; Advanced detection tools for real-time mitigation of misinformation and bias; Interdisciplinary collaborations among computer scientists, ethicists, and domain experts to develop context-aware, ethically sound multimodal systems.

By systematically mapping domains, benefits, and challenges, this review not only clarifies the current state of multimodal AI research but also lays a clear agenda for future investigations aimed at enhancing methodological rigor, ethical governance, and broad-scale applicability.

CRediT (Contributor Roles Taxonomy)

Contribuciones de los autores: Conceptualización: JD **BR**; Metodología: NY **RC**; Investigación: JD **BR**, NY **RC**; Redacción y preparación del borrador original: JD **BR**; Redacción, revisión y edición: NY **RC**; Análisis formal: JD **BR**; Adquisición de fondos: No aplica.

Financiamiento: Los autores declaran que no se requirió de adquisición de fondos para este estudio.

Declaración de disponibilidad de datos: Los datos se encuentran en el artículo.

Agradecimientos: Los autores desean expresar su agradecimiento a las plataformas de Inteligencia Artificial generativa, Gemini y Grok, por haber brindado asistencia en la etapa de pulido del lenguaje del documento, así como en la generación de la Figura 2. Es importante destacar que la investigación, el análisis formal y la interpretación de los resultados presentados son responsabilidad exclusiva de los autores humanos (JD **BR** y NY **RC**).

Conflicto de interés: Los autores declaran que no existe conflicto de interés.

REFERENCIAS

[1] G. Joshi, R. Walambe, K. Kotecha, "A Review on Explainability in Multimodal Deep Neural Nets," *IEEE Access*, vol. 9, pp. 59800–59821, 2021, doi: 10.1109/ACCESS.2021.3070212

[2] M. Chen, *et al*., "Replanteamiento de la moderación de contenido multimodal desde un ángulo asimétrico con modalidad mixta," *ArXiv*, vol. abs/2305.10547, 2023, doi: 10.48550/arXiv.2305.10547

[3] M. J. Page *et* al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[4] CASP, "CASP Qualitative Checklist: 10 questions to help you make sense of qualitative research (Version 6)." 2018.

[5] D. P. Serrano, J. P. M. Belda, P. M. Ramallal, "Impacto y oportunidades de la inteligencia artificial en el arte visual: personalización de la creatividad con Stable Diffusion y Controlnet," in *Manipulación en imágenes visuales y sonoras en ficción y no ficción*, Dykinson, 2023, pp. 894–909.

[6] D. Palenzuela López, *Herramientas de IA generativa para crear puentes literarios: exploración visual con La Casa de Bernarda Alba*, 2024.

[7] F. Guerrero-Solé, C. Ballester, "El impacto de la Inteligencia Artificial Generativa en la disciplina de la comunicación," Hipertext.net, no. 26, pp. 1–9, 2023.

[8] J. L. Martín Prada, *et al*., *La creación artística visual frente a los retos de la inteligencia artificial. Automatización creativa y cuestionamientos éticos*, 2024.

[9] O. Cordón García, "Inteligencia Artificial en Educación Superior: Oportunidades y Riesgos," *Revista Interuniversitaria de Investigación en Tecnología Educativa*, pp. 16-27, Dec. 2023, doi: 10.6018/riite.591581

[10] M. H. Sarkhoosh, S. Gautam, C. Midoglu, S. Sabet, P. Halvorsen, "Multimodal AI-Based Summarization and Storytelling for Soccer on Social Media," in *Proceedings of the 15th ACM Multimedia Systems Conference*, 2024, pp. 485–491.

[11] V. Liu, "Beyond text-to-image: Multimodal prompts to explore generative AI," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–6.

[12] M. R. Morris, J. R. Brubaker, "Generative ghosts: Anticipating benefits and risks of AI afterlives," *arXiv* preprint arXiv: 2402.01662, 2024.

[13] A. De Bruyn, V. Viswanathan, Y. S. Beh, J. K.-U. Brock, F. Von Wangenheim, "Artificial intelligence and marketing: Pitfalls and opportunities," *Journal of Interactive Marketing*, vol. 51, no. 1, pp. 91–105, 2020.

[14] Y. K. Dwivedi *et al*., "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management*, vol. 57, p. 101994, 2021.

[15] A. Fitriawijaya, T. Jeng, "Integrating multimodal generative ai and blockchain for enhancing generative design in the early phase of architectural design process," *Buildings*, vol. 14, no. 8, p. 2533, 2024.

[16] Y. Zhou, H. J. Park, "Una aplicación multimodal aumentada con IA para esbozar diseños conceptuales," *International Journal of Architectural Computing*, vol. 21, pp. 565–580, 2023, doi: 10.1177/14780771221147605

[17] C. G. Cabay Calderon, "Integración de la Inteligencia Artificial con otras tecnologías en la creación decontenidos para redes sociales," B.S. thesis, Universidad Tècnica de Ambato, Facultad de Jurisprudencia y Ciencias 2023.

[18] M. Jin, *et al*., "Enfoque multimodal para analizar grandes datos de redes sociales y medios de comunicación," *Discourse, Context and Media*, vol. 40, p. 100467, 2021, doi: 10.1016/J.DCM.2021.100467

[19] Y. Lu, *et al*., "Ética e IA: una revisión sistemática sobre preocupaciones éticas y estrategias relacionadas para el diseño con IA en el ámbito sanitario," AI, 2022, doi: 10.3390/ai4010003

[20] L. Sun, *et al*., "Una encuesta exhaustiva sobre el contenido generado por IA (AIGC): una historia de la IA generativa desde GAN hasta ChatGPT," ArXiv, vol. abs/2303.04226, 2023, doi: 10.48550/arXiv.2303.04226.

[21] S. Dadgar, M. Neshat, "A novel hybrid multi-modal deep learning for detecting hashtag incongruity on social media," *Sensors*, vol. 22, no. 24, p. 9870, 2022.

[22] J. Dana, V. Agnus, F. Ouhmich, B. Gallix, "Multimodality imaging and artificial intelligence for tumor characterization: current status and future perspective," in *Seminars in Nuclear Medicine*, Elsevier, 2020, pp. 541–548.

Científica, vol. 29, no. 2, pp. 01-11, July-December 2025,
ISSN 2594-2921, Instituto Politécnico Nacional, MÉXICO
e290208 | DOI: 10.46842/ipn.cien.v29n2a08

[23] M. Imran, F. Ofli, D. Caragea, A. Torralba, "Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions," *Information Processing & Management*, vol. 57, no. 5, p. 102261, Sep. 2020, doi: 10.1016/j.ipm.2020.102261

[24] B. Jaltotage, J. Lu, G. Dwivedi, "Use of Artificial Intelligence Including Multimodal Systems to Improve the Management of Cardiovascular Disease," *Canadian Journal of Cardiology*, 2024.

[25] M. Tong, *et al.*, "Detección de anomalías mediante aprendizaje profundo multimodal," vol. 12941, pp. 129414V-129414V–7, 2023, doi: 10.1117/12.3011769

[26] I. Jebadurai, A. M. Vedamanickam, P. U. Kumar, "Design of Generative Multimodal AI Agents to Enable Persons with Learning Disability," Companion Publication of the *25th International Conference on Multimodal Interaction*, 2023, doi: 10.1145/3610661.3617514.

[27] T. K. Mohd, N. Nguyen, A. Y. Javaid, "Multi-modal data fusion in enhancing human-machine interaction for robotic applications: a survey," *arXiv* preprint arXiv:2202.07732, 2022.

[28] P. J. Standen *et al*., "An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities," *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1748–1765, 2020.

[29] T. Baltrušaitis, C. Ahuja, L. P. Morency, "Challenges and applications in multimodal machine learning," *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*, pp. 17–48, 2021.

[30] H. Alwahaby, M. Cukurova, Z. Papamitsiou, M. Giannakos, "The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review," *The Multimodal Learning Analytics Handbook*, pp. 289–325, 2022.

Científica, vol. 29, no. 2, pp. 01-11, July-December 2025,
ISSN 2594-2921, Instituto Politécnico Nacional, México
e290208 | DOI: 10.46842/ipn.cien.v29n2a08