e-JURNAL PENYELIDIKAN DAN INOVASI
ISLAMIC INNOVATION CENTER FOR TECHNOLOGY, UIS
https://ejpi.uis.edu.my/index.php/ejpi
e-ISSN 2289-7909

VOL. 13 NO. 2 (AUGUST, 2026) Special Edition AIDI
Page 260-273
DOI: https://doi.org/10.53840/e-jpi.v13i2.432

Received: 17 July 2026 Accepted: 18 August 2026 Published: 31 August 2026

RESEARCH TRENDS IN AI-DRIVEN VIDEO PRODUCTION: A SYSTEMATIC BIBLIOMETRIC INVESTIGATION

Mus'ab Khalili Bin Rosli[1], Muhammad Farhan Fauzan Bin Masaat[2], Nur Muizz Bin Mohamed[3], Mohd Khairul Mustakim Bin Mustapa Kamal[4], Ahmad Khalid Bin Kamarul Saman[5]

Faculty of Creative Multimedia & Computing
Universiti Islam Selangor
Email: [1]m.khalili@uis.edu.my, [2]farhanfauzan@uis.edu.my, [3]muizz@uis.edu.my, [4]khairulmustakim@uis.edu.my, [5]ahmadkhalid@uis.edu.my

ABSTRACT

Artificial intelligence (AI) has fundamentally reshaped the world of video production, moving from simple automated edits to complex generative models capable of creating realistic content from scratch (Xie et al., 2025). This study maps the intellectual evolution of this field from 2016 to 2026, focusing on key publication trends and thematic shifts. A systematic bibliometric study of 2,371 peer-reviewed papers included in the Scopus database was carried out. The metrics included the total number of publications (TP), total citations (TC), the number of cited publications (NCP), average citations per publication (C/P), and average citations per cited publication (C/CP). The data reveals a massive surge in research output starting in 2021, with total publications reaching a peak of 796 in 2025. While volume is currently at an all-time high, the highest citation impact per paper occurred in 2019, indicating that the foundational frameworks of the field were established during that time. Thematically, research is clustered into four main areas: core technical architectures, media synthesis, deepfake detection and human ethics. This investigation provides a comprehensive roadmap for researchers navigating the fast growth of AI-driven media production.

Keywords: *Bibliometric Analysis; AI Video Production; Deepfake; Generative AI; Industry 5.0.*

INTRODUCTION

Over the past century, there has been a significant shift in the development of picture and video technology. The use of Digital picture Processing (DIP) in scientific domains to greatly improve picture quality in the 1960s marked a turning point in this history. For example, in 1964, the American Jet Propulsion Laboratory (JPL) made history by processing photos of the Moon using digital technology. This innovation significantly improved the field's technological capabilities (Sarfraz, 2020; Hidayat et al., 2025). This led to the development of an era marked by intelligent and autonomous image and later video content production (Luo, Hao, & Tong, 2018). In the context of AI-Driven Video Production, artificial intelligence (AI) has progressed beyond its traditional role as a tool for automation and optimization, emerging as a

collaborative partner in creative processes (Gurel, 2026; Cruz Oña & Andrade Sánchez, 2025). Recent advances in AI-driven video production have enabled the generation of high-quality visual content from text, image, and audio inputs, a phenomenon widely recognized as AI-Generated Content (AIGC) (Xie et al., 2025; Cao et al., 2023). These developments have expanded creative possibilities, enhanced production efficiency, and reduced operational costs across media industries (Hinojosa Becerra et al., 2024; Franco Lazarte, 2025). However, the rapid pace of technological advancement has also created challenges for scholars and practitioners seeking to understand the evolving intellectual landscape of this field. Consequently, systematic academic investigations are necessary to map research developments, identify emerging trends, and assess the broader implications of AI-driven video production technologies.

PROBLEM STATEMENT

The central challenge in this field is that while AI-driven video technology is evolving rapidly. The academic research documenting it has become increasingly fragmented and difficult to manage (Xie et al., 2025). Rapid breakthroughs which transitioning from early Generative Adversarial Networks (GANs) to contemporary Diffusion Models and Transformer-based architectures like Sora—have occurred within such a narrow window that existing reviews struggle to provide a holistic view. Consequently, new researchers find it nearly impossible to distinguish between papers that establish truly foundational methodologies and those that merely capitalize on current industry "hype" (Franco Lazarte, 2025). Without a long-term, longitudinal perspective to organize these shifts, the scholarly community lacks a "big picture" understanding of how the domain has transitioned from basic computer vision tasks to complex AI-Generated Content (AIGC).

Furthermore, there is a distinct lack of clarity regarding the actual research impact versus the sheer volume of scientific output being produced. While the Total Number of Publications (TP) has increased exponentially, reaching a recorded peak of 796 articles in 2025, the data suggests that high productivity does not always equate to high influence. For instance, while volume is at an all-time high, the highest Average Citations per Publication (C/P) actually occurred in 2019, indicating that seminal works are being buried under the recent surge of output. This gap makes it difficult to measure scientific progress accurately, as the academic community lacks a bibliometric overview that distinguishes between general output (TP) and truly influential Number of Cited Publications (NCP).

In the context of Industry 5.0, which emphasizes human-centricity and ethical design, the current scholarly landscape is also "curiously silent" on the broader social and ethical implications of these tools. While technical clusters around media synthesis (red cluster) and deep learning (green cluster) are expanding, there is a lag in research addressing deepfake misuse, data privacy, and the erosion of digital authenticity (Vizoso et al., 2021; Faranguillo, 2023). Policymakers and creative professionals are currently forced to navigate these disruptive technologies without a structured, data-driven roadmap that identifies key institutional hubs and emerging thematic hotspots (Kaleel & Alomari, 2024). This study directly addresses these issues by providing a systematic bibliometric framework to organize the scattered landscape of AI-driven video production research.

LITERATURE REVIEW

The academic trajectory of AI-driven video production has transitioned from a theoretical niche to a central pillar of the Artificial Intelligence Generated Content (AIGC) movement (Xie et al., 2025; Cao et al., 2023).

Historically, research in this domain was "latent," focusing primarily on rule-based computer vision tasks like object recognition. However, the landscape shifted dramatically around 2014 with the introduction of Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and Variational Autoencoders (VAEs) (Kingma & Welling, 2013), which allowed for the first realistic synthesis of visual data. This era established the foundational ability to create synthetic media, though early models often struggled with temporal coherence and high-resolution stability.

The real "turning point" for the industry occurred between 2020 and 2021 (Cruz Oña & Andrade Sánchez, 2025). During this period, the field entered an intensive expansion phase characterized by the emergence of Diffusion Models and Transformer-based architectures. Unlike their predecessors, these models—most notably OpenAI's Sora and many more demonstrated the capacity to act as "world simulators," generating complex, minute-level videos from diverse inputs including text, images, and audio (Xie et al., 2025; Zhu et al., 2024). This technological leap has revolutionized the audiovisual industry, enabling unprecedented efficiency in scriptwriting, automated editing, and the creation of visual effects (Franco Lazarte, 2025). A high-profile example of this convergence can be seen in the film *The Irishman*, which utilized deep learning to "rejuvenate" actors across different historical eras (Botha & Pieterse, 2020; Cruz Oña & Andrade Sánchez, 2025).

As we enter the era of Industry 5.0, the research focus is evolving from purely technical milestones toward a human-centric paradigm (Gurel, 2026; European Commission, 2025a). AI is no longer viewed just as an automation tool but as a creative partner that supports ideation and co-creation (Grech et al., 2023). However, this rapid democratization of technology has created a "double-edged sword" effect. While the creative possibilities are vast, they are accompanied by significant risks, including the proliferation of deepfakes, misinformation, and intellectual property concerns (Vizoso et al., 2021; Kietzmann et al., 2020).

Consequently, current scholarly trends are increasingly dominated by digital forensics and the development of "Ethics by Design" frameworks (Gurel, 2026). Research is now tasked with balancing visual fidelity with information integrity, ensuring that AI-driven production enhances human storytelling without compromising digital authenticity or cultural depth (Kaleel & Alomari, 2024). This investigation maps these multidimensional shifts, providing a structured roadmap of how the field has grown from basic image processing to a sophisticated, interdisciplinary domain.

METHODOLOGY

This study adopts a systematic bibliometric approach to analyze the trend of AI-driven video production. To ensure academic accuracy, this study followed the PRISMA guidelines for identifying and screening relevant literature. In keeping with the bibliometric standard, this approach promoted credibility and reproducibility by making data extraction and analysis consistent (Page et al., 2021; Hidayat et al., 2025).

Data Source and Search Strategy

The metadata for this investigation was retrieved from the Scopus database on 9 June 2026. Scopus was selected for its comprehensive coverage of engineering, computer science, and multimedia journals. The search field was restricted to Article Titles to ensure high relevance to the core topic. To ensure the precision and integrity of the final corpus, a comprehensive manual screening was conducted on the initial 2,371 identified records. This phase involved a meticulous review of the titles, keywords, and abstracts of each document to systematically identify and exclude any potential false positives. Following this rigorous manual assessment, it was determined that all 2,371 records strictly aligned with the study's stringent inclusion criteria and thematic focus, justifying their retention for the final systematic bibliometric investigation.The following complex Boolean search string was applied to capture the intersection of AI technologies and the creative video industries by using the advance search in Scopus database:

TITLE-ABS-KEY (("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Generative AI" OR "Computer Vision" OR "Neural Network" OR "Image Processing") AND ("Multimedia" OR "Video*" OR "Film*" OR "Animation" OR "Digital Content" OR "Digital Media" OR "Creative Industr*" OR "Broadcasting" OR "Computer Graphics" OR "Visual Effects")) AND PUBYEAR > 2015 AND PUBYEAR < 2027 AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))*

Inclusion and Exclusion Criteria

To focus the analysis on contemporary trends and high-quality peer-reviewed work, several stringent filters were applied during the selection process. First, the study's time frame was restricted to publications between 2016 and 2026 to capture the most rapid period of growth and innovation in AI technologies concentrating on the Trends in AI-Driven Video Production.

However, while the research timeframe encompasses publications up to 2026, it is important to clarify that the data for the final year is partial. Since the metadata was extracted from the Scopus database on 9 June 2026, the records for this year only reflect the scholarly output up to the first six months and do not represent the total cumulative publications for the entire year of 2026.

Second, only articles written in the English language were included to ensure linguistic consistency across the dataset. Furthermore, to maintain a focus on original empirical work, the search was limited strictly to journal articles, while other document types such as books, conference proceedings, and reviews were excluded.

Finally, only articles that had reached their "final" publication stage were considered, ensuring that the metadata and citation counts used for the bibliometric analysis were stable and complete.

Data Cleaning and Screening

The initial search and screening process identified a total of n = 2,371 records. After a thorough review of the metadata, it was determined that zero (0) records needed to be removed as they all met the stringent inclusion criteria. Therefore, the final dataset included for bibliometric analysis and thematic mapping was n = 2,371 articles.

FINDINGS AND DISCUSSIONS

This chapter presents the comprehensive findings of the systematic bibliometric investigation into AI-driven video production research conducted between 2016 and 2026. The analysis is structured into two primary dimensions: first, a quantitative performance analysis documenting the field's growth and citation influence; and second, a qualitative thematic mapping of the intellectual structure using network visualization. By synthesizing these metrics, this section provides a data-driven roadmap that identifies key research trajectories, important milestones, and the evolving creative-technical landscape of the Industry 5.0 era.

Temporal Growth and Citation Impact

To map the evolutionary trajectory of AI-driven video production, it is essential to analyze both the volume of annual scientific productivity and the resulting scholarly influence within the academic community. This section examines the growth of the field through five key bibliometric indicators: Total Publications (TP), Number of Cited Publications (NCP), Total Citations (TC), Average Citations per Publication (C/P), and Average Citations per Cited Publication (C/CP). Assessing these metrics over a decadal timeframe allows for the identification of foundational periods—where truly influential methodologies were established—versus phases of rapid "hype" or industry expansion.

Table 1: Year of Publication (2016-2026)*

Year	TP	%	NCP	TC	C/P	C/CP	h	g
2026	508	21.43%	92	201	0.40	2.18	6	9
2025	796	33.57%	590	3930	4.94	6.66	26	37
2024	472	19.91%	442	7698	16.31	17.42	42	58
2023	259	10.92%	252	9073	35.03	36.00	49	82
2022	168	7.09%	163	7211	42.92	44.24	49	79
2021	102	4.30%	102	4776	46.82	46.82	41	66
2020	39	1.64%	38	2629	67.41	69.18	21	39
2019	11	0.46%	10	871	79.18	87.10	6	11
2018	6	0.25%	6	162	27.00	27.00	5	6
2017	5	0.21%	5	88	17.60	17.60	3	5
2016	5	0.21%	5	36	7.20	7.20	4	5

**Data as of mid-year 2026*

Analysis of the bibliometric performance data reveals that the field of AI-driven video production research reached its absolute peak of scientific productivity in 2025, recording 796 articles (TP), which accounts for 33.57% of the entire dataset over the ten-year period. A significant growth shift was detected starting in 2021, when research output surged to 102 articles, representing a sharp increase compared to the mere 39 articles recorded in 2020. While publication volume peaked in 2025, the highest scholarly impact in terms of Total Citations (TC) was achieved earlier in 2023 with a record of 9,073 citations. Quality indicators further support this trend: the h-index reached its maximum of 49 between 2022 and 2023, while the highest g-index was recorded in 2023 with a value of 82. Additionally, the average influence per paper was found to be highest during the early phase, with Average Citations per Publication (C/P) and per Cited Publication (C/CP) reaching an absolute peak in 2019 at 79.18 and 87.10, respectively.

An in-depth analysis of these trends offers several critical inferences regarding the intellectual evolution of the field. First, the exceptionally high citation rates in 2019, despite a very small number of publications (TP = 11), indicate that the articles published that year represent the foundational frameworks (seminal

works) that set the direction for the entire decade. The exponential growth observed between 2021 and 2025 reflects a phase of massive expansion driven by the emergence of revolutionary technologies such as Diffusion Models and world-simulating architectures like Sora.

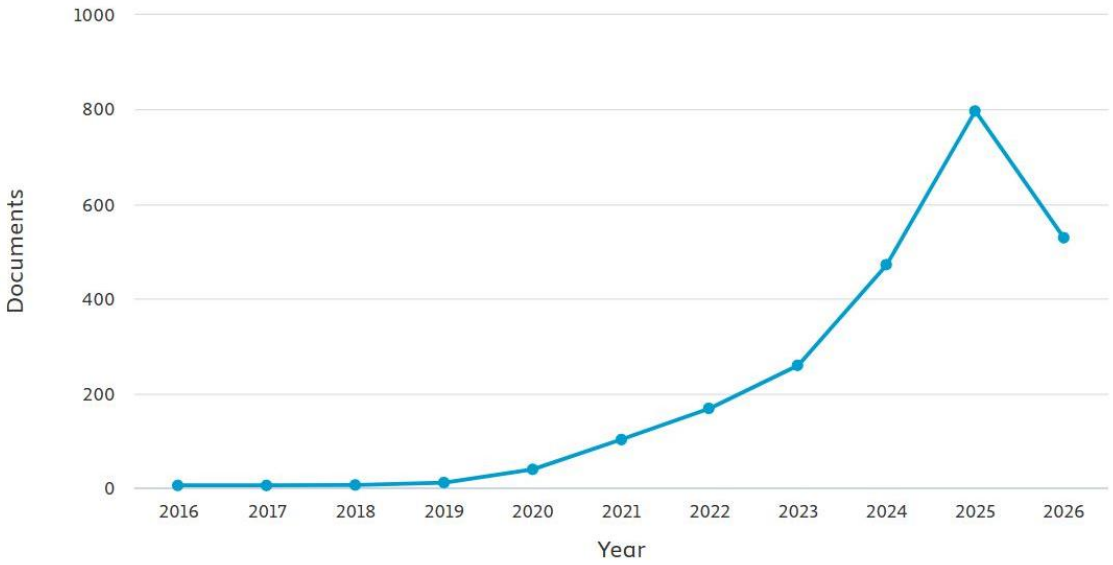


Figure 1: Document by Year of Publication (2016-2026) *

**Data as of mid-year 2026*

As can be seen in the figure 1, the notable decline in publication in 2026 and citation values for 2025-2026 (such as a C/P of 0.40 in 2026) should not be interpreted as a decline in research quality; rather, it illustrates the citation lag phenomenon, where the most recent articles require a longer timeframe to be discovered and cited by other researchers.

It must also be emphasized that the data for 2026 is incomplete, as it only includes records up to June 9, 2026, suggesting that this year has the possibility to break existing productivity records once a full year of data is collected. The stability of high h-index values in 2022 and 2023 also proves that the field has reached a level of intellectual maturity, characterized by a core group of researchers and works that exert a consistent and significant influence on the global academic community.

In conclusion, this data shows a shift from a specialized technical field (2016–2020) to a significant worldwide academic priority (2021–2026), with the early foundational studies having the greatest qualitative influence.

Thematic Mapping and Intellectual Structure: An Analytical Investigation

To fully understand the area, it's important to examine its conceptual landscape and intellectual structure, beyond only its quantitative performance. Keyword co-occurrence analysis gives a visual picture of how important concepts are interrelated, indicating significant research "hotspots" in AI-driven video creation over the previous decade. The qualitative method identifies patterns and themes that mere publication counts cannot.

The VOSviewer program was used to create a network representation of the final dataset of 2,371 articles. This visualization depicts author keywords as nodes, with size proportional to frequency of occurrence. The lines linking the nodes show co-occurrence, with the proximity between terms implying a higher level of academic relationship.

Table 2: Top Keywords

Author Keywords	Total Publications (TP)	Percentage (%)
Deepfake	708	29.86%
Deepfake Detection	580	24.46%
Deep Learning	531	22.40%
Face Recognition	302	12.74%
Artificial Intelligence	284	11.98%
Feature Extraction	269	11.35%
Human	238	10.04%
Deepfakes	213	8.98%
Fake Detection	204	8.60%
Digital Forensics	178	7.51%
Learning Systems	178	7.51%
Convolutional Neural Networks	177	7.47%
Generative Adversarial Networks	172	7.25%
Videorecording	152	6.41%
Benchmarking	151	6.37%
Semantics	143	6.03%
Forgery Detections	137	5.78%
Detection Methods	135	5.69%
Diffusion Model	135	5.69%
Video Recording	132	5.57%

Based on the quantitative data shown in Table 2, it provides a clear reference and picture of the current research goals and conceptual framework of AI-driven video production. Direct findings from the keyword frequency analysis show a high domination of digital integrity concerns, with "Deepfake" and "Deepfake Detection" are the most often used terms, accounting for 29.86% and 24.46% of the 2,371 articles, respectively.

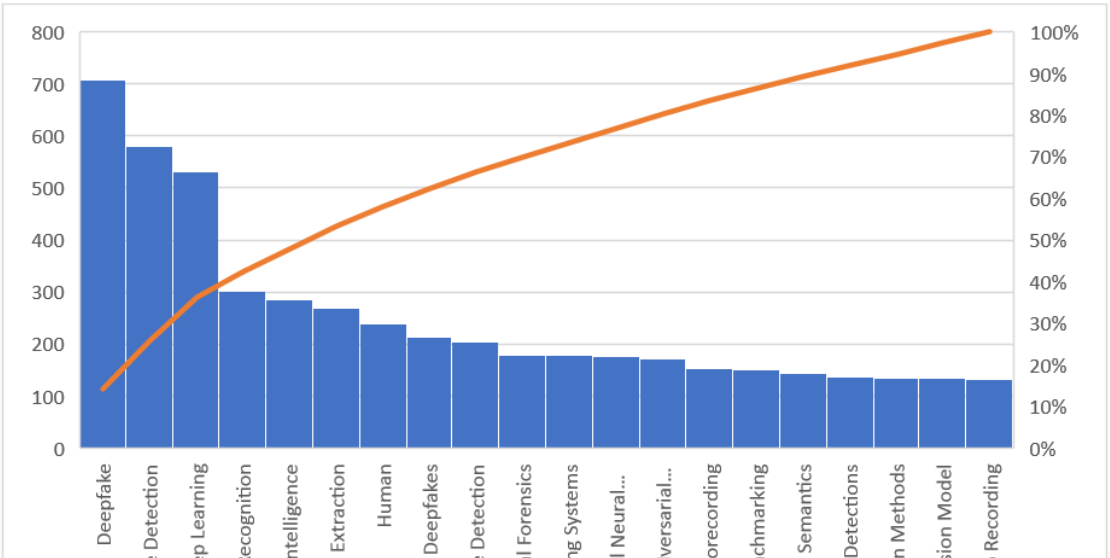


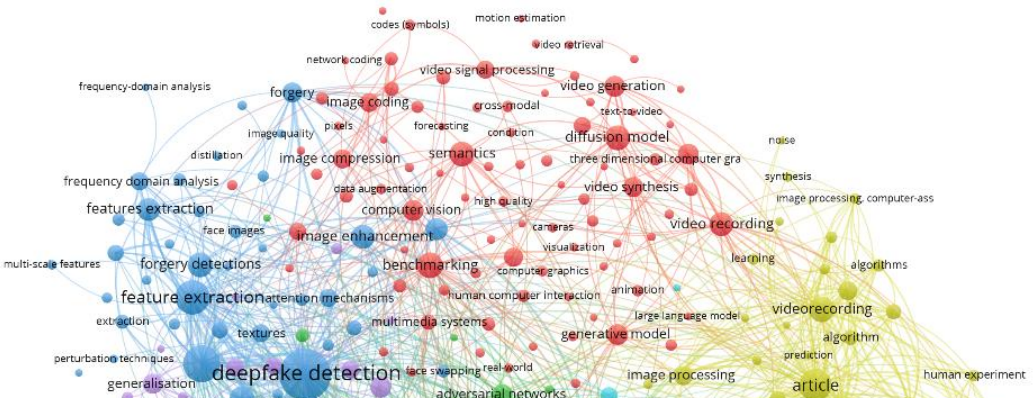
Figure 2: Top Keywords Chart

As seen in the chart (Figure 2), these top keywords, together with "Deep Learning" (22.40%), account for more than half of the field's total thematic focus. This concentration suggests that the scholarly community places a high priority on synthetic media and its verification. Other important methodological pillars include "Face Recognition" (12.74%), "Artificial Intelligence" (11.98%), and "Feature Extraction" (11.35%). Furthermore, technical structures such as "convolutional neural networks" and "generative adversarial networks" remain core components of the literature, while the advent of "diffusion models" among the top 20 illustrates the current trend toward more sophisticated generative architectures.

A closer examination of the data in Table 2 and Figure 2 reveals numerous significant themes about the field's history. The steep starting slope of the cumulative percentage line in Figure 2 indicates a "winner-takes-all" theme distribution, with "Deepfake" being the major lens for examining AI video output. This implies that the industry's creative advancement is predominantly perceived through a defensive and security-focused lens.

The high frequency of creation-focused phrases and verification-focused procedures suggests a technical "arms race". As generative realism advances, it requires prompt academic reaction in Digital Forensics (7.51%) and Forgery Detection (5.78%). The constant ranking of "Deep Learning" and "CNNs" demonstrates that the domain remains basic, with advances in creative synthesis tied to key machine learning achievements. The presence of "Human" as a top-ten term (10.04%) indicates a shift in research towards Industry 5.0, which focuses on user effect and behavioral reactions rather than only automation. This transition implies that the subject is developing from a merely technical study to a complex, multidisciplinary arena in which technology advancement is balanced by ethical responsibility and human narrative.

Based on the data from Table 2, and Chart on Figure 2 its highlights several key facts about the current state of the field, most notably the dominance of digital integrity issues. The keywords "Deepfake" and "Deepfake Detection" are found to be the most frequent, appearing in 708 and 580 publications respectively, which represents over 54% of the total dataset when combined.



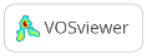


Figure 3: Network Visualization Map for 'Keyword' from VOS Viewer

Based on Figure 3 above, it presents the Network Visualization Map for Keyword co-occurrence network generated using VOSviewer. The network visualization map reveals the intellectual structure of AI-driven multimedia and video-related research through the relationships among frequently occurring author keywords. The size of each node indicates keyword frequency, while the links represent co-occurrence relationships. Different colors denote thematic clusters that collectively define the major research directions within the field.

The domain is physically separated into four different color-coded clusters, where Green Cluster (Technical Backbone), which is dominated by fundamental approaches like Deep Learning (TP=531) and Convolutional Neural Networks (TP=177), the Red Cluster (Media Synthesis), which is focused on creation techniques like Diffusion Models and Video Synthesis, the Blue Cluster (Security and Integrity), which is focused on defense through Feature Extraction and Digital Forensics; and the Yellow Cluster (Social and Ethical Impact), which addresses human-centric variables like Ethics, Misinformation, and Human Experiments.

In the network visualization map above, the "Deep Learning" node acts as the primary bridge, showing the highest number of links to nodes in every other cluster, thereby proving its central position in the field's intellectual structure, indicating that research in AI-driven multimedia and video technologies has evolved around interconnected areas of deepfake detection, deep learning, video generation, and human-centered AI applications.

A dynamic research environment where technical innovation is constantly balanced by the need for digital integrity and ethical monitoring is shown by mapping the conceptual landscape of AI-driven video creation using a co-occurrence analysis of 2,371 articles. The resultant network visualization shows how the discipline has evolved from simple, rule-based computer vision tasks to a sophisticated, multi-modal generative environment. The domain's present trajectory and future possibilities are defined by the four key thematic clusters that make up this progression which is:

The Technical Backbone (Green Cluster): The Methodological Engine

The Technical Backbone (Green Cluster) from the Network Visualization Map (figure 3) serving as the methodological engine for the entire domain, this cluster exhibits the highest degree of centrality within the map (Shaheen et al., 2023). The analysis confirms that Deep Learning (TP=531) acts as the primary bridge node, connecting creative synthesis with defensive security. Analytically, the dominance of foundational algorithms specifically Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) demonstrates that current trends in video creation are a direct outcome of small but significant advances in fundamental machine learning (Hidayat et al., 2025). In the past, this backbone supplied the infrastructure required for pattern recognition, facilitating the transition to the advanced, self-governing learning systems that characterize the modern age.

Beyond its role as a computational engine, the Green Cluster functions as a critical conceptual bridge between the creative output of the Red Cluster (Media Synthesis) and the ethical imperatives of the Yellow Cluster (Social and Ethical Impact) while breakthroughs in core machine learning (Green) enable the high-fidelity, world-simulating capabilities of AIGC (Red), they also dictate the architectural constraints under which ethical monitoring occurs. Specifically, the technical backbone provides the necessary frameworks for Explainable AI (XAI) and 'Ethics by Design,' which are essential for addressing the biases and lack of transparency identified in the human-centric paradigm. By serving as this link, the Green Cluster ensures that the pursuit of generative realism (Red) is systematically tethered to the responsibility of safeguarding human dignity and cultural integrity (Yellow), transforming raw technical capability into a balanced tool for Industry 5.0.

Media Synthesis (Red Cluster): From Simple Editing to World Simulation

The study's creative core is represented by this cluster, which shows a dramatic transition from simple video editing to comprehensive Artificial Intelligence Generated Content (AIGC). The emergence of Diffusion Models and architectures such as OpenAI's Sora is the most disruptive trend seen in this bibliometric analysis (Xie et al., 2025). With the use of these technologies, artificial intelligence has evolved from a simple automation tool to a "world simulator" that can create tales that make sense from text or visual inputs without the need for conventional filmmaking. The Media Synthesis cluster highlights the field's rapid growth and its significance in transforming the audiovisual industry by democratizing high-fidelity visual storytelling.

Security and Integrity (Blue Cluster): The Digital Authenticity Arms Race

The enormous dominance of the "Deepfake" and "Deepfake Detection" keywords, which together account for more than 54% of the whole dataset, is a crucial conclusion of this study. From an analytical perspective, this cluster is a defensive academic reaction to the Media Synthesis cluster's quick developments. A technical "arms race" has resulted from this relationship: as generative realism advances, the academic community quickly reacts with increasingly complex feature extraction and digital forensics techniques to validate material (Xie et al., 2025; Malik et al., 2022). This suggests that in an era of manufactured disinformation, the main academic concern is not just how to "produce" video but also how to guarantee its validity.

Social and Ethical Impact (Yellow Cluster): The Transition to Industry 5.0

The essence of this last cluster, which prioritizes human-centricity and ethical design in line with the Industry 5.0 standard, is a reflection of the field's intellectual maturity. A thorough examination shows that

the "curious silence" about ethics observed in previous stages of AIGC development is finally being addressed by researchers. This cluster examines the psychological and social effects of synthetic media on consumers by concentrating on human experimentation and AI ethics. The prominence of these issues implies that the future of AI-driven video creation will be assessed based on the effective integration of "Ethics by Design" frameworks to safeguard cultural integrity and human dignity rather than just visual accuracy (Gurel, 2026; Kaleel & Alomari, 2024).

The bibliometric data clearly demonstrates that Deep Learning and Deepfakes are the primary themes of contemporary study. However, the mapping implicitly points to a highly interwoven ecosystem in which technical proficiency, creative production, and ethical responsibility are intrinsically intertwined. This connection offers a thorough road map for the subject, demonstrating how it has developed from a limited technical discipline into a sophisticated, multidisciplinary field with human narrative as its primary focus.

CONCLUSION

This systematic bibliometric investigation successfully mapped the intellectual evolution of AI-driven video production over a transformative decade (2016–2026). The findings indicate that the field has evolved from a relatively niche technical domain into a major global academic priority, with a substantial increase in scientific output beginning in 2021. Although 2025 recorded the highest level of productivity with 796 published articles, the impact analysis revealed that 2019 represented the most critical phase, during which several foundational and influential works established the trajectory of contemporary technological developments.

From a thematic perspective, the study identified four interconnected research clusters: the technical “engine room” (Green Cluster), the creative frontier of media synthesis (Red Cluster), digital integrity and security (Blue Cluster), and the ethical and human-centered dimension (Yellow Cluster). One of the most significant findings is the dominance of Deepfake-related research, which accounts for more than 54% of the overall research focus. This highlights the existence of an ongoing “arms race” between advances in generative realism and developments in digital forensics. The field is also transitioning from a phase primarily focused on automation towards the Industry 5.0 era, where AI is no longer viewed solely as a technical tool but increasingly as a creative collaborator operating within clearly defined ethical boundaries (Gurel, 2026; European Commission, 2025a).

By offering a clear road map for navigating an increasingly fragmented AI research field, this paper provides a significant contribution. The discovery that Deep Learning is a crucial link between digital security and creative media synthesis provides scholars with important guidance for their future study. The findings about the transition to human-centered design highlight the critical necessity for industry practitioners and policymakers to incorporate a "Ethics by Design" paradigm at every step of generative media development. To guarantee that digital storytelling in the future is based on authenticity, legitimacy, and cultural integrity, such a strategy is crucial.

For developers, this paradigm translates into the implementation of concrete technical safeguards, such as digital watermarking of synthetic datasets to ensure content provenance and the adoption of transparent algorithmic reporting through Explainable AI (XAI) to proactively identify and mitigate embedded biases. By integrating these actionable protocols, the industry can ensure that the pursuit of generative realism remains systematically tethered to the responsibility of safeguarding human dignity and cultural integrity.

Although offering thorough insights into the topic, it is important to recognize a few limits. First, certain breakthroughs published through other channels, such as highly specialized conference proceedings and grey literature, may have been overlooked because the Scopus database was the only source of information used. Second, productivity estimates for 2026 should be viewed as suggestive rather than reflective of the entire yearly production because data for that year were only available until June 2026. Lastly, significant research viewpoints from areas where English is not the predominant language of academic communication may have been overlooked due to the emphasis only on English-language publications.

In the future, the research community is urged to investigate more expansive interdisciplinary aspects and transcend strictly technical bounds. Further research on the long-term psychological consequences of AI-generated material on public perception is desperately needed, as is the creation of more fair intellectual property frameworks for joint works between humans and AI. Future studies should concentrate on enhancing algorithmic transparency and making sure that technologies like Sora and diffusion-based models strive for both social responsibility and visual quality. AI-driven video creation has the potential to become a catalyst for more inclusive, meaningful, and human-centered storytelling by striking a balance between ethical integrity and technical innovation.

REFERENCES

Botha, J., & Pieterse, H. (2020). Fake News And Deepfakes: A Dangerous Threat For 21st Century Information Security. Proceedings Of The 15th International Conference On Cyber Warfare And Security (ICCWS), 57–66.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A Comprehensive Survey Of AI-Generated Content (AIGC): A History Of Generative AI From GAN To Chatgpt. arXiv. https://arxiv.org/abs/2303.04226

Cruz Oña, A. O., & Andrade Sánchez, E. (2025). Convergence Between Artificial Intelligence And Audiovisual Production: A Bibliometric Analysis Of Its Evolution, Trends And Contributions (2000–2024). Visual Review, 17(5), 153–171. https://doi.org/10.62161/revvisual.v17.5811

European Commission. (2025a). What Is Industry 5.0? https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

Faranguillo, J. (2023). La Inteligencia Artificial Generativa Y Su Impacto En La Creación De Contenidos Mediáticos. Methaodos Revista De Ciencias Sociales, 11(2). http://dx.doi.org/10.17502/mrcs.v11i2.710

Franco Lazarte, E. G. (2025). Audiovisual Narrative In The Age Of Artificial Intelligence: Advances, Trends And Challenges: A Systematic Review. Journal Of Educational And Social Research, 1–15.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems, 27.

Grech, A., Mehnen, J., & Wodehouse, A. (2023). An extended AI-experience: Industry 5.0 In Creative Product Innovation. Sensors, 23(6), 3009. https://doi.org/10.3390/s23063009

Gurel, E. (2026). AI-Driven Experiences In Cultural And Creative Industries: A Review Of Literature And Development Of A Multifaceted Framework. The Service Industries Journal, 46(7–8), 583–622. https://doi.org/10.1080/02642069.2025.2542822

Hidayat, E. Y., Hastuti, K., & Muda, A. K. (2025). Artificial Intelligence In Digital Image Processing: A Bibliometric Analysis. Intelligent Systems With Applications, 25, 200466. https://doi.org/10.1016/j.iswa.2024.200466

Hinojosa Becerra, M., Marín Gutiérrez, I., & Maldonado Espinosa, M. (2024). Inteligencia Artificial Y La Producción Audiovisual. Comunicación Social Ediciones Y Publicaciones.

Kaleel, A., & Alomari, M. S. (2024). Integrating Artificial Intelligence In Public Relations And Media: A Bibliometric Analysis Of Emerging Trends And Influences. Iraqi Journal For Computer Science And Mathematics, 5(1), 13–24. https://doi.org/10.52866/ijcsm.2024.05.01.002

Kietzmann, J., Lee, L., McCarthy, I., & Kietzmann, T. (2020). Deepfakes: Trick Or Treat? Business Horizons, 63(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. Arxiv. https://arxiv.org/abs/1312.6114

Luo, C., Hao, Y., & Tong, Z. (2018). Research On Digital Image Processing Technology And Its Application. In Proceedings Of The 2018 8th International Conference On Management, Education And Information (MEICI 2018) (pp. 587–592).

Malik, A., Kuribayashi, M., Abdullahi, S., & Khan, A. (2022). Deepfake Detection For Human Face Images And Videos: A Survey. IEEE Access, 10, 18757–18775. https://doi.org/10.1109/ACCESS.2022.3151184

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline For Reporting Systematic Reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Sarfraz, M. (2020). Introductory Chapter: On Digital Image Processing. In Digital Imaging (Pp. 1–9). Intechopen. https://doi.org/10.5772/intechopen.92277

Shaheen, H., Ravikumar, K., Lakshmipathi Anantha, N., Uma Shankar Kumar, A., Jayapandian, N., & Kirubakaran, S. (2023). An Efficient Classification Of Cirrhosis Liver Disease Using Hybrid Convolutional Neural Network–Capsule Network. Biomedical Signal Processing And Control, 80, 104152. https://doi.org/10.1016/j.bspc.2022.104152

Vizoso, Á., Vaz-Álvarez, M., & López-García, X. (2021). Fighting Deepfakes: Media And Internet Giants' Converging And Diverging Strategies Against Hi-Tech Misinformation. Media And Communication, 9(1), 291–300. https://doi.org/10.17645/mac.v9i1.3494

Xie, W., Hu, A., Xie, Q., Chen, J., Wan, R., & Liu, Y. (2025). Bibliometric Analysis And Review Of AI Based Video Generation: Research Dynamics And Application Trends (2020–2025). Discover Computing, 28, 130. https://doi.org/10.1007/s10791-025-09628-9

Zhu, Z., Wang, X., Zhao, W., et al. (2024). Is Sora A World Simulator? A Comprehensive Survey On General World Models And Beyond. arXiv. https://arxiv.org/abs/2405.03520

Zupic, I., & Čater, T. (2015). Bibliometric Methods In Management And Organization. Organizational Research Methods, 18(3), 429–472. https://doi.org/10.1177/1094428114562629