West Science Interdisciplinary Studies
Vol. 03, No. 11, November and 2025: pp. 2098-2111

Brand Safety in the Deepfake Era: A Bibliometric Analysis

Loso Judijanto
IPOSS Jakarta, Indonesia and losojudijantobumn@gmail.com

ABSTRACT

This paper examines the evolution of scholarship on brand safety in the deepfake age by a bibliometric analysis of publications at the convergence of deepfakes, artificial intelligence, cybersecurity, and marketing. Utilizing a prominent citation database, we extracted and refined pertinent articles, reviews, and conference papers published from 2010 to 2025, then analyzing them with Bibliometrix and VOSviewer. Performance metrics, keyword co-occurrence, co-authorship, affiliations, and international collaboration networks were employed to delineate the intellectual and social framework of the discipline. The findings indicate two primary streams: a technical-security stream concentrating on deep learning-based detection and cyber threat intelligence, and an applied stream focusing on AI-driven commerce, metaverse environments, and consumer reactions to deepfake material. India and a limited number of partnering universities emerge as crucial knowledge centers, but other regions remain comparatively isolated. The article presents an ecosystem perspective on brand safety concerning synthetic media, highlights significant research deficiencies, and delineates avenues for future theoretical and empirical investigations

Keywords: *Brand Safety, Deepfake, Synthetic Media, Artificial Intelligence, Deep Learning, Cybersecurity, Metaverse, Bibliometric Analysis.*

1. INTRODUCTION

Brand safety has emerged as a critical issue in modern marketing, as firms increasingly depend on programmatic advertising, social media, and user-generated content to engage customers. Brand safety, in its conventional definition, pertains to safeguarding a brand's reputation by regulating the contexts and types of content with which its advertisements are connected, therefore prevent associations with hate speech, violence, disinformation, or other detrimental material [1], [2]. Misaligned associations can undermine trust, impair brand equity, and provoke customer reaction or regulatory examination [3]. The advent of synthetic media and generative arti ficial intelligence has broadened the brand safety agenda from simple placement control to encompass the integrity and authenticity of content, especially as malicious entities can create highly realistic representations of brands and their representatives [4], [5].

Deepfake technologies—AI-generated audio, images, and video that accurately replicate real individuals—have progressed swiftly in the past decade, facilitated by advancements in deep learning, enhanced computing capabilities, and extensive datasets obtained from digital platforms [6], [7]. Originally emerging as online curiosities and amusement memes, deepfakes are increasingly acknowledged as instruments for disinformation, harassment, fraud, and reputational damage aimed against persons, institutions, and corporations [6], [7]. Researchers caution that synthetic media exacerbate "truth decay," obscure the distinction between genuine and fake conten t, and create new opportunities for exploitation in political and commercial domains [5]. These dynamics create significant risks for brands whose value proposition relies significantly on consistency, credibility, and perceived authenticity.

For marketers and brand managers, deepfakes present a category of hazards that are more intrinsic and challenging to handle than conventional adjacency issues. A faked video of a CEO making derogatory comments, a counterfeit product recall statement, or a synthetic influencer end

orsement can disseminate swiftly before official rectifications reach the public, particularly in dynamic social media contexts [6], [7]. In contrast to traditional brand safety concerns that can typ ically be addressed by blocklists, whitelists, and collaborations with verification providers, deepfake risks directly distort the brand and its representatives as the content itself, rather than merely altering the context surrounding an advertisement. This advances brand governance towards more refined digital risk management, in corporating AI-driven detection tools, specific contractual protections with artists and platforms, and crisis communication protocols especially tailored for synthetic media occurrences [1], [8]. A further layer of complication emerges from the phenomenon known as the "liar's dividend," which involves the tactical exploitation of public understanding of deepfakes to refute the legitimacy of genuine yet detrimental content. Studies in political communication indicate that politicians may manipulate suspicion about digital media by asserting that authentic recordings are deepfakes, thereby undermining accountability procedures [8]. A like trend may be observed in branding: companies encountering authentic scandals may be inclined to attribute responsibility to "AI manipulation," while stakeholders find it challenging to differentiate between genuine brand communications and counterfeit messages [7], [8]. Consequentl y, deepfakes generate misleading content regarding brands and undermine the overarching epi stemic context in which brand assertions and narratives are assessed [6].

Concurrently, practitioner reports and industry standards for brand safety and "brand suitability" are increasing, yet the scholarly literature at the nexus of brand safety and deepfakes remains disjointed. Digital advertising and platform governance mostly address concerns such ad fraud, viewability, and generic content dangers, while deepfake research is predominantly influenced by computer science, law, security studies, and political communication [5], [6]. A limited number of research specifically investigate how synthetic media transform brand safety policies, influence consumer perceptions of authenticity, or modify the mechanisms by which brand confidence can be compromised and reinstated. Concurrently, industry discussions are evolving f rom limited avoidance-based frameworks of brand safety to more sophisticated concepts of appropriateness and contextual alignment; nevertheless, comprehensive scholarly analysis of these transitions in the deepfake era remains limited [2], [4]. This establishes a disparity between rapidly evolving industry practices and comparatively fragmented academic ideas.

Notwithstanding the increasing awareness of deepfake-related hazards in industry and policy discussions, the academic literature remains deficient in a cohesive, data-driven analysis of the conceptualization and examination of brand safety within the realm of synthetic media. The connection between technological research on deepfake production and detection and marketing-focused efforts on brand reputation, customer trust, and crisis management is still constrained [5], [6]. Empirical research directly investigating consumer reactions to deepfake occurrences involving brands, the efficacy of corporate response plans, or the enduring effects on brand equity is scarce. Current evaluations generally discuss deepfakes or digital brand safety in a broad context, lacking a rigorous analysis and quantification of the precise overlap between these areas. This fragmentation hinders researchers and practitioners from comprehending the structure, evolution, and deficienci es of the research domain on brand safety in the deepfake era as a unified body of knowledge [9].

This study intends to perform a thorough bibliometric analysis of international research concerning "Brand Safety in the Deepfake Era" to overcome these deficiencies. The study has three primary objectives: (1) to chart the progression of scientific publications at the nexus of brand safety, deepfakes, and synthetic media, encompassing temporal trends and prominent authors, institutions,

countries, and journals; (2) to discern prevailing themes, conceptual clusters, and theoretical frameworks through keyword co-occurrence, co-citation, and co-authorship analyses; and (3) to identify research deficiencies and suggest future research trajectories that can enhance both academic exploration and managerial strategies in addressing brand safety risks in light of advancing deepfake technologies [9]. This study use known bibliometric methods to deliver a systematic, evidence-based analysis of the academic community's response to brand safety challenges in the deepfake era and identifies areas for urgent further investigation.

2. METHODS

This paper employs a bibliometric analysis to systematically delineate the scholarly landscape at the convergence of brand safety, deepfakes, and synthetic media. In accordance with the recognized protocols for bibliometric research [9], a prominent multidisciplinary citation database, such as Scopus or Web of Science, was utilized as the principal data source due to its extensive coverage of peer-reviewed journals and comprehensive citation metadata. The search strategy integrated keywords pertaining to brand safety and brand suitability (e.g., “brand safety”, “brand suitability”, “advertising safety”, “ad adjacency risk”) with terminology associated with deepfake and synthetic media (e.g., “deepfake”, “synthetic media”, “AI-generated video”, “synthetic advertising”, “generative AI” AND “brand”). The search was limited to English publications, con ference papers, and reviews within a specified timeframe (e.g., 2010–2025) to encompass the era w hen deepfake technology emerged in academic discussions. Only papers classified as final or in-press peer-reviewed outputs were maintained; non-academic materials such as editorials, news articles, and notes were eliminated [9].

Subsequent to the first retrieval, the dataset underwent cleaning and standardization before analysis. Duplicate records from search strings were eliminated, and bibliographic details (authors' names, affiliations, titles, abstracts, keywords, and references) were verified for consistency. Author names and institutional affiliations were standardized to rectify spelling discrepancies and different iate common names, while keywords were normalized by consolidating synonyms and alternative spellings (e.g., “deepfake” vs. “deep fake”; “brand safety” vs. “advertising safety”) to enhance the reliability of co-occurrence analyses. The final dataset was exported in a compatible format (e.g., BibTeX or CSV) and analyzed using Bibliometrix in R and its Biblioshiny web interface for perform ance metrics (publication trends, most prolific authors, institutions, countries, and sources), alongside VOSviewer for scientific mapping visualizations, including co-authorship, co-citation, and keyword co-occurrence networks [10], [11]

The analytical approach integrated descriptive bibliometric metrics with network and thematic analysis to fulfill the research objectives. Descriptive metrics, including annual publication growth, citation counts, prominent journals and authors, and national productivity, were employed to delineate the temporal evolution and structural attributes of the discipline [9]. Network studies of co-authorship and institutional collaboration elucidated the social structure of the research community, whereas co-citation and keyword co-occurrence networks were employed to discern conceptual clusters and prevailing themes pertinent to brand safety and deepfakes. Cluster solu tions and density visualizations produced in VOSviewer were analyzed to identify principal researc h streams (e.g., technological deepfake detection, platform governance, consumer trust and authenticity, brand risk and crisis management) and to emphasize underexplored intersections (Aria & Cuccurullo, 2017; van Eck & Waltman, 2010). During the process, methodological selections and parameter configurations (including minimum thresholds for citations, co-authorship, or keyword frequency) were recorded to guarantee transparency and replicability, while recognizing intrinsic limitations in database coverage, English-language bias, and the swiftly changing landscape of research on generative AI and deepfakes.

3. RESULTS AND DISCUSSION

3.1 Network Visualization

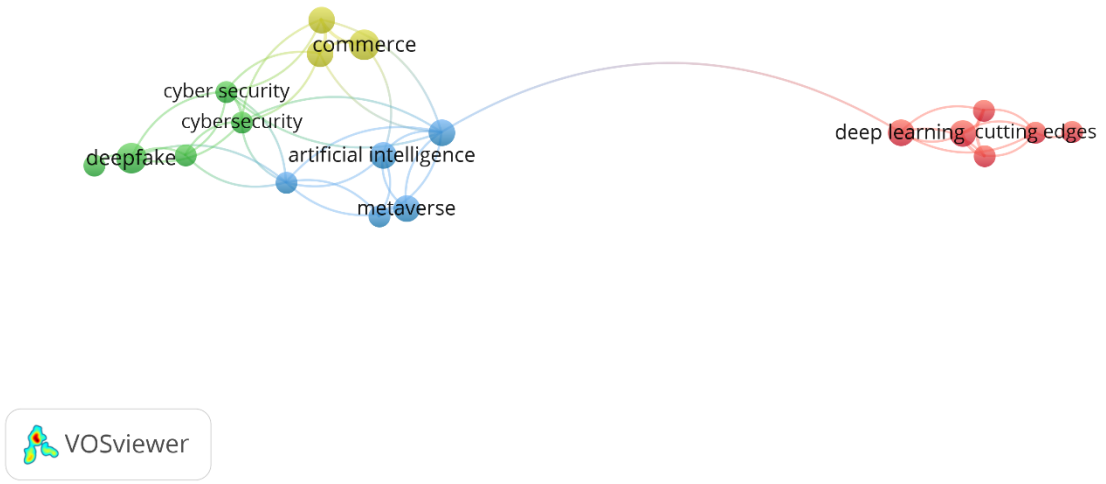

Figure 1. Network Visualization
Source: Data Analysis Result, 2025

The VOSviewer map depicts illustrates three primary term clusters that organize the research landscape. On the left, there exists a greenish cluster encompassing "deepfake," "cyber security," and "cybersecurity"; in the center, a blue-yellow area featuring "artificial intelligence," "metaverse," and "commerce"; and on the far right, a red technical cluster surrounding "deep learning" and "cutting edges." The dimensions of the nodes signify their comparative significance (occurrence/links), while the colored clusters denote that these phrases frequently co-occur in the same publications, thereby establishing rather cohesive theme communities.

The green cluster including deepfake and cybersecurity indicates a body of research that positions deepfakes predominantly as a security and risk issue. Currently, deepfakes are primarily associated with dangers such cyber-attacks, impersonation, fraud, and the erosion of digital trust, rather than branding or marketing. The tight association between deepfake and the dual spellings of cybersecurity suggests frequent co-occurrence, indicating that authors commonly address deepfakes alongside broader cybersecurity issues, such data breaches, identity theft, or information integrity.

The center blue-yellow zone including artificial intelligence, the metaverse, and commerce signifies a more application-focused, business, and market viewpoint. Artificial intelligence serves as a central hub, linking commercial applications, such e-commerce and digital business models, with developing environments like the metaverse. This indicates an expanding corpus of research focused on AI-driven commerce within immersive or virtual environments, where user experience, digital transactions, and platform governance are paramount, and where deepfakes and brand safety may arise as significant issues, even if not explicitly stated.

The red cluster, using deep learning and advanced technologies, is visually distinct from the others on the right, linked solely by a slender connection to the central AI node. This signifies a more technical, methods-oriented study domain, concentrated on algorithmic innovations and cutting-edge deep learning methodologies. These studies likely pertain to architectures,

performance benchmarks, and model developments, rather than commerce, the metaverse, or brand-level ramifications. The tenuous connection indicates that while deep learning forms the foundation of deepfakes and artificial intelligence broadly, the technical literature remains mostly isolated from practical discourse concerning commerce, cybersecurity, and media contexts.

The overall organization indicates a disparity and a potential: deepfake and cybersecurity efforts are grouped together but lack robust integration with the commerce/metaverse/brand-relevant domain, while the avant-garde deep learning initiatives are conceptually even more remote. This map visually substantiates the assertion that the literatures on technical, security, and business/commerce aspects are only tenuously interconnected in a study on brand safety during the deepfake era. It indicates that forthcoming research may serve as a conduit—connecting deep learning and deepfake detection to tangible commercial and metaverse environments, while amalgamating cybersecurity considerations with brand safety, consumer trust, and platform governance.

3.2 Overlay Visualization

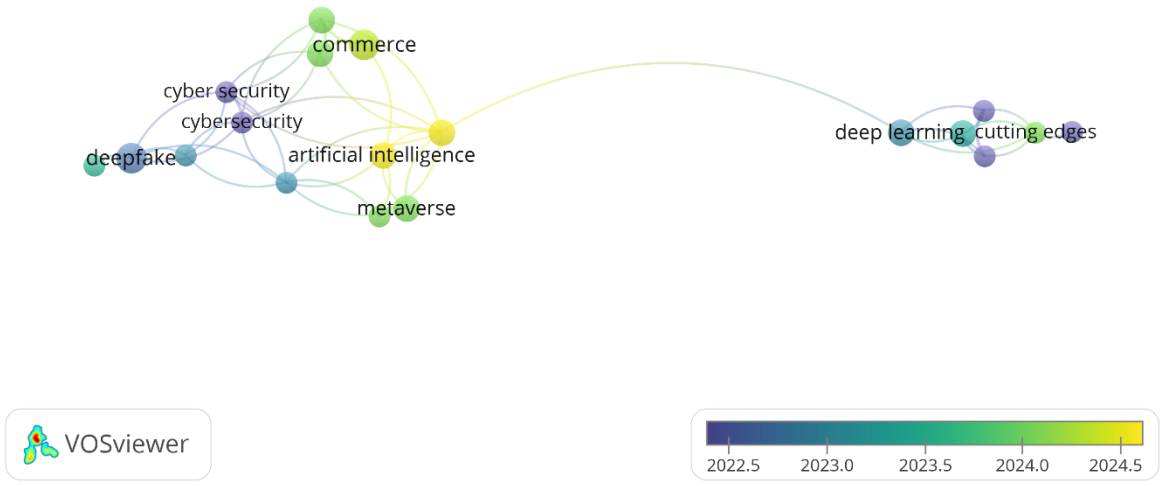

Figure 2. Overlay Visualization
Source: Data Analysis Result, 2025

The overlay visualization illustrates incorporates a time dimension to the keyword network by color-coding each node based on its average publication year, ranging from dark blue (about 2022) to yellow (approximately 2024). On the left, phrases such as “deepfake,” “cyber security,” and “cybersecurity” are presented in darker blue-teal hues, signifying their earlier introduction into the dataset. This indicates that the initial phase of research in this corpus characterized deepfakes predominantly as a security and risk problem, positioned within discussions of cybersecurity and information integrity. Deepfake emerges as a recognized, albeit nascent, focal point associated with safeguarding digital infrastructures rather than directly pertaining to branding or consumer-oriented business.

As one approaches the center, the terms "artificial intelligence," "metaverse," and "commerce" exhibit lighter green to yellow hues, indicating a more recent focus. AI is now seen not merely as a technological foundation but increasingly as a facilitating tool for commercial

applications and immersive environments. The trinity of metaverse, commerce, and AI suggests that researchers have lately started investigating the intersection of AI-driven systems, virtual environments, and digital marketplaces—an area where concerns such as identity, trust, and the possibility for deepfake-enabled manipulation of experiences or transactions are particularly pertinent. This temporal change signifies a transition from exclusively protective cybersecurity narratives to talks focused on business, platforms, and markets, where brand safety issues begin to emerge.

On the far right, the "deep learning" and "cutting edges" cluster is rendered in a teal-to-purple gradient, emerging slightly later than the initial cyber/deepfake nodes yet remaining distinct, indicative of a modern technological frontier. These phrases denote method-centric endeavors that advance the forefront of algorithms and architectures; they are linked to the broader framework via artificial intelligence yet remain mostly distinct from the commerce, metaverse, and security terminologies. The combined overlay pattern indicates a developing domain: commencing with security-focused deepfake and cybersecurity initiatives, subsequently broadening to encompass AI-driven commerce and metaverse environments, while concurrent technical research in deep learning advances fairly independently. Your research on "Brand Safety in the Deepfake Era" underscores the emergence of brand-related implications in the contemporary core of the field, highlighting the necessity for more robust conceptual connections between advanced technological developments, cybersecurity principles, and the evolving commercial and branding discussions.

3.3 Citation Analysis

This study initially analyzed the most frequently referenced publications in the dataset to ascertain the fundamental intellectual underpinnings of research on deepfakes, cybersecurity, and brand-related results. These extensively referenced studies encompass the technical advancements in deepfake detection as well as the behavioral and marketing ramifications of synthetic media, including cyber threat intelligence, political communication, consumer engagement, brand credibility, and loyalty. Collectively, they offer a succinct overview of the progression in literature from algorithmic methods for identifying manipulated content to a more nuanced comprehension of how deepfake videos and advertising influence consumer perceptions, values, and brand affiliations. The principal attributes of the most-cited documents are encapsulated in Table 1.

Table 1. The Most Impactful Literatures

Citations	Authors and year	Title
55	Amerini, I., Caldelli, R. (2020)	Exploiting Prediction Error Inconsistencies through LSTM-based Classifiers to Detect Deepfake Videos
39	Saxena, R., Gayathri, E. (2021)	Cyber threat intelligence challenges: Leveraging blockchain intelligence with possible solution
29	Sharma, I., Jain, K., Behl, A., ... Giannakis, M., Dwivedi, Y. (2023)	Examining the motivations of sharing political deepfake videos: the role of political brand hate and moral consciousness
9	Salam, M.A., Rayun, S.M.N., Islam, W., ... Firmansyah, E.A., Kalinaki, K. (2024)	Consumer engagement: Exploring deepfake applications in consumer marketing communication
8	Liu, M., Wang, J., Qian, X., Li, H. (2024)	Audio-Visual Temporal Forgery Detection Using Embedding-Level Fusion and Multi-Dimensional Contrastive Loss

Citations	Authors and year	Title
8	Shin, J.-Y., Suk, J., Chung, J.-E. (2023)	Consumer Responses to Fashion in the Metaverse: A Text-Mining Analysis on Online News Comments
6	Sardana, F., Mishra, K.K., Singh, A., Saini, N. (2024)	Transforming social media marketing through deepfake technology
4	Dimuthu Maduranga Arachchi, H.A., Samarasinghe, G.D. (2024)	Impact of Deepfake Advertising Attributes on Consumers' Hedonic & Utilitarian Values and Brand Credibility
4	Jellali, A., Fredj, I.B., Ouni, K. (2023)	Data Augmentation for Convolutional Neural Network DeepFake Image Detection
3	Arachchi, H.A.D.M., Samarasinghe, G.D., Wickramasinghe, A. (2025)	Seeing is Believing: Exploring Deepfake Video Ads and Brand Loyalty in the Experience Economy

Source: Scopus, 2025

Table 1 illustrates that the most significant contributions can be categorized into two complimentary streams. The initial category is a technical-security stream, exemplified by research from Amerini and [12], [13], [14], [15], which formulate sophisticated detection models or investigate cyber threat intelligence and data augmentation techniques to alleviate deepfake threats. The seco nd stream pertains to consumer marketing, encompassing the works of [16], [17], [18], [19], and two studies by Arachchi and associates (2024, 2025). These studies investigate the motivations behind sharing political deepfakes, the application of deepfakes in consumer marketing, responses to fashion within the metaverse, and the effects of deepfake advertising on hedonic and utilitarian val ues, brand credibility, and consumer loyalty. The coexistence of these streams underscores a gradual transition from perceiving deepfakes solely as a security threat to acknowledging them as a potent, albeit hazardous, instrument in digital marketing and brand communication specifically the nexus that drives the current bibliometric analysis on brand safety in the deepfake era.

3.4 Density Visualization

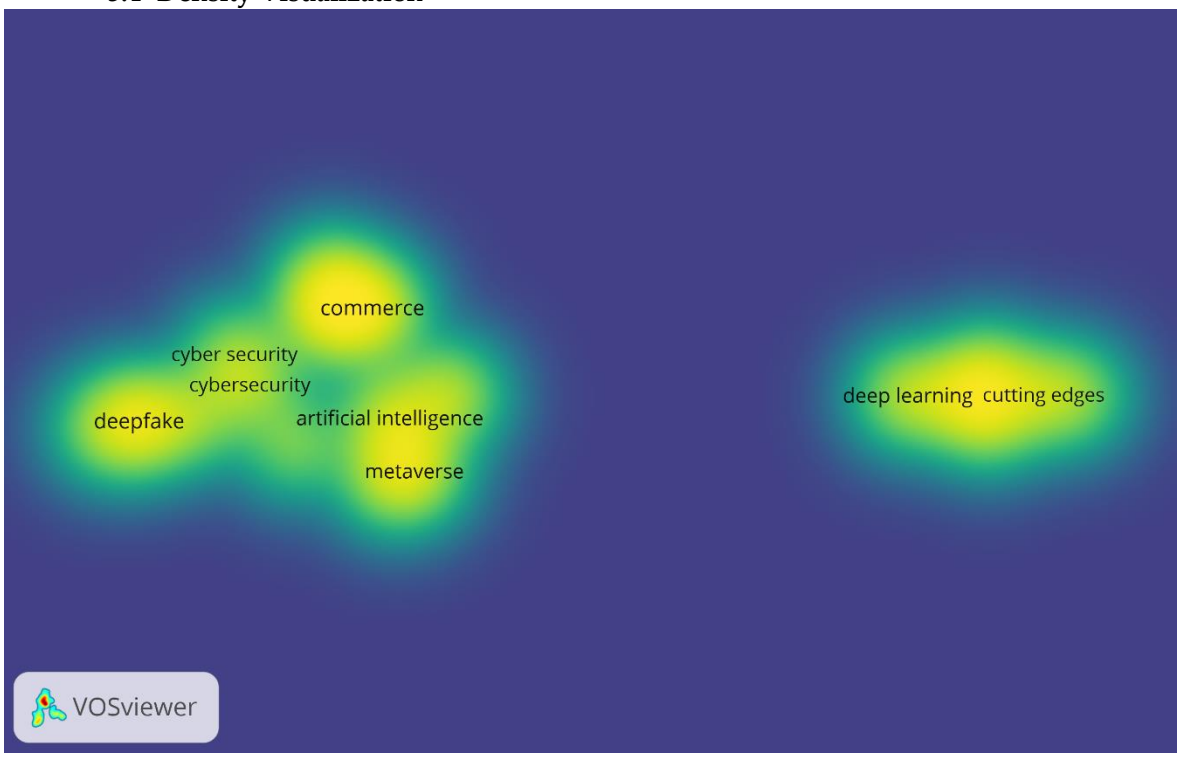

Figure 3. Density Visualization

Source: Data Analysis Result, 2025

The density graphic illustrates the areas of most research activity based on term co-occurrence. The vibrant yellow-green region on the left signifies a concentrated cluster of publications pertaining to the terms "commerce," "artificial intelligence," "metaverse," "cyber security/cybersecurity," and "deepfake." The predominant focus is on business and artificial intelligence, indicating that a significant portion of the papers in your collection associates AI with commercial or marketplace contexts, frequently encompassing metaverse environments and contextualized by cybersecurity or deepfake threats. The focal point of the topic is not solely technical; it resides at the convergence of AI-driven business applications, virtual environments, and security/privacy issues.

A distinct high-density zone emerges on the right, centered around "deep learning" and "cutting edges," signifying a robust albeit somewhat insular technological domain dedicated to advanced algorithms and model development. The spatial separation between this cluster and the commerce/AI/deepfake domain suggests that numerous advanced deep learning research endeavors remain inadequately interconnected with commerce, metaverse, or brand-related concerns. The discourse on Brand Safety in the Deepfake Era reveals a bifurcation in the literature: (1) an applied stream addressing AI, commerce, metaverse, cybersecurity, and deepfakes, and (2) a technical stream focused on deep learning methodologies—underscoring a distinct opportunity to integrate sophisticated detection techniques with tangible brand safety and marketing applications.

3.5 Co-Authorship Network

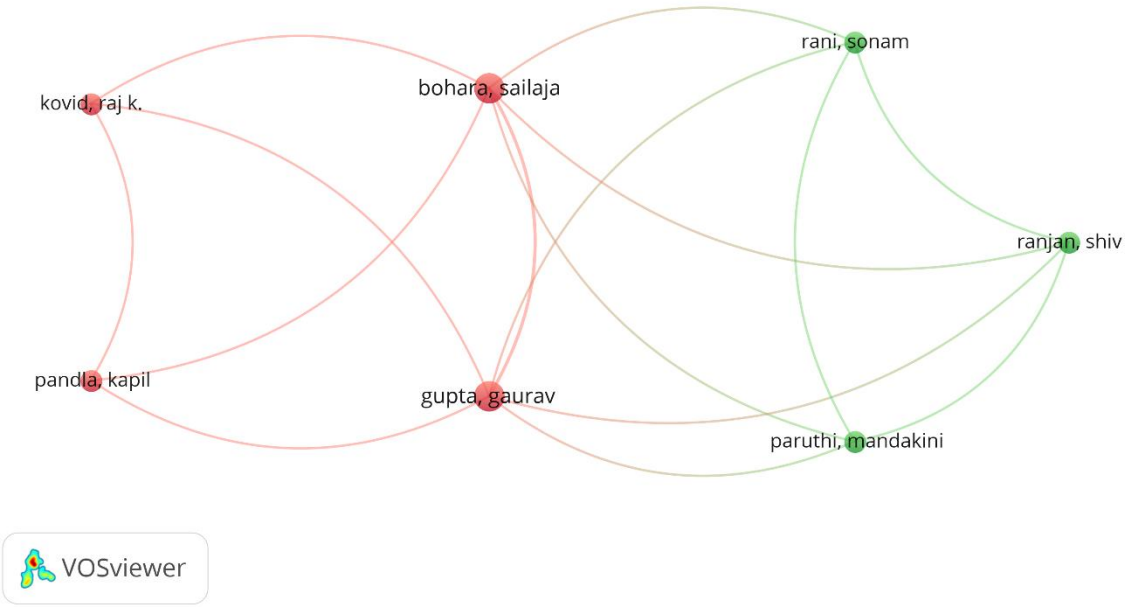

Figure 4. Author Visualization
Source: Data Analysis Result, 2025

The VOSviewer co-authorship network visualization a co-authorship network, with each node symbolizing an author and the connections denoting collaborative publications. The graphic illustrates two closely linked subgroups: on the left, kovid, raj k., pandla, kapil, bohara, sailaja, and gupta, gaurav constitute a red cluster that engages in intensive collaboration, indicating a solid core team or a persistent research relationship. To the right, the green cluster, consisting of Rani, Sonam, Ranjan, Shiv, and Paruthi, has comparable internal cohesiveness with Mandakini. The authors Bohara, Sailaja, and Gupta play a pivotal role connecting the red and green groups, serving as bridging authors who co-publish with members of both clusters. This framework suggests that knowledge and methodologies are likely disseminated throughout the network via these key links, while the peripheral writers depend on them to attain more collaboration prospects and exposure within the discipline.

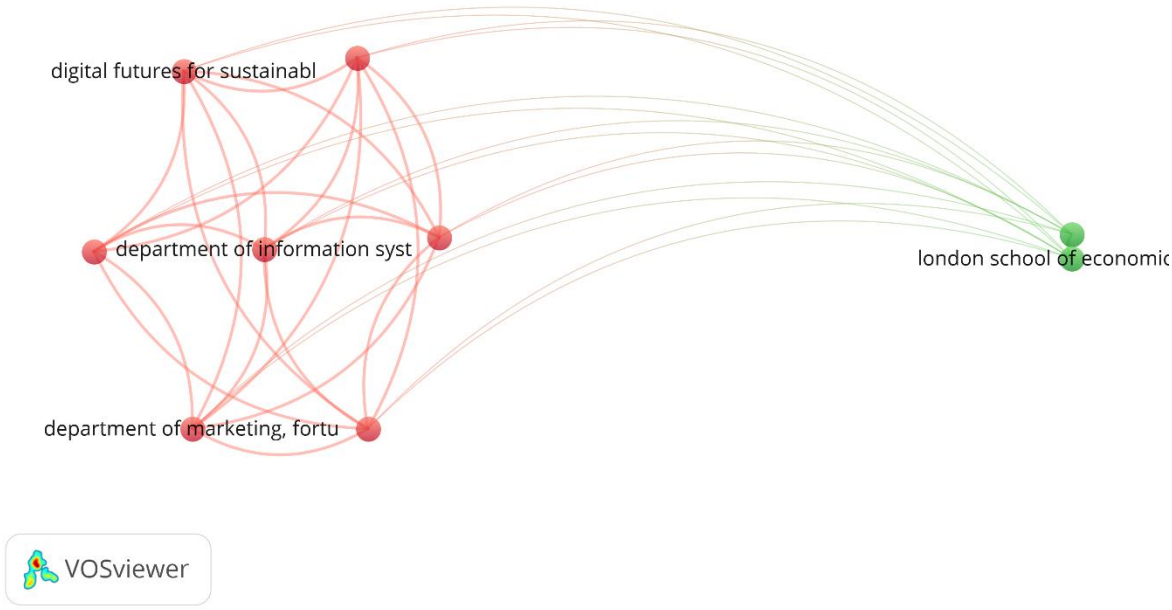

Figure 5. Affiliation Visualization
Source: Data Analysis Result, 2025

This VOSviewer map illustrates an institutional collaboration network, with each node denoting an affiliation and the lines representing co-authored papers among them. The red cluster on the left consists of strongly linked entities, including the marketing department, the information systems department, "digital futures for sustainable," and an additional internal node, signifying a robust structure of collaboration within the same university or research institution. These units provide a highly cohesive core that regularly co-publishes on subjects pertaining to digital futures, marketing, and information systems. The London School of Economics is represented as a separate green node on the right, interconnected by numerous edges to all constituents of the red cluster. This pattern indicates that LSE functions as a crucial external partner, participating in multiple collaborative projects with the central institution while maintaining its status as a distinct collaboration cluster, rather than integrating into the internal network. The map illustrates a unified internal research ecosystem that is externally focused, with LSE acting as a prominent international partner in this deepfake, AI, and commerce-related knowledge network.

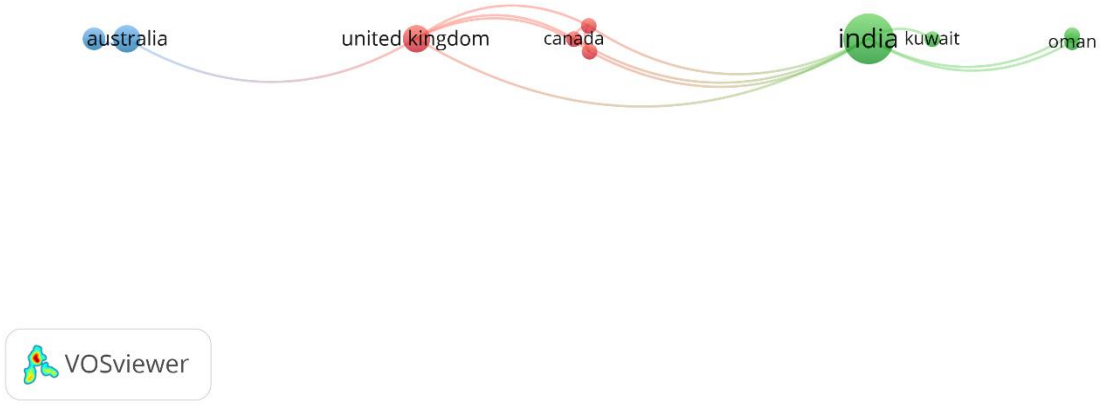

Figure 6. Country Visualization
Source: Data Analysis Result, 2025

The VOSviewer co-country collaboration map depicts the collaborative network across countries in your dataset. The size of each node indicates the publishing volume of each country, while the links denote co-authored publications between countries. India unequivocally dominates the network as the principal node, serving as the central hub that engages extensively with Canada, the United Kingdom, Kuwait, and Oman, suggesting that a significant portion of the research on deepfakes, AI, and commerce within this corpus is propelled by Indian scholars collaborating with both Western and Gulf counterparts. Kuwait and Oman manifest as minor yet interconnected nodes mostly linked to India, indicating the emergence of regional partnerships. On the western side, Canada and the United Kingdom are of considerable size and interconnected, representing a secondary collaboration hub between these nations that also links back to India. Australia occupies a more periphery position with a tenuous connection to the broader network, indicating limited yet present engagement. The map illustrates a region physically situated in India, exhibiting significant yet uneven linkages to Anglophone and Gulf nations.

Discussions

Practical Implications

This study provides multiple practical recommendations for professionals addressing brand safety in the age of deepfakes. The maps indicate that research, and consequently practical technologies, is predominantly focused in two distinct areas: technical deepfake detection and AI-driven commerce, metaverse, and marketing. This indicates to brand managers the necessity of establishing cross-functional teams that integrate IT/cybersecurity with marketing, legal, and business communication, rather than regarding deepfake detection as only a technological enhancement. Secondly, the significance of buzzwords like commerce, artificial intelligence, and metaverse indicates that brand safety measures must be broadened beyond traditional adplacement regulations to encompass virtual environments, influencer networks, and AI-generated content streams. Companies must formulate explicit regulations for the permissible and impermissible uses of synthetic media in campaigns, establish verification and removal methods with platforms, and

investment in monitoring tools to track both brand-generated and user-generated deepfake material. Third, the country and affiliation networks underscore India and a limited number of institutions as knowledge hubs; practitioners—particularly regulators, industry associations, and global brands—can utilize this map to pinpoint potential knowledge partners for training, collaborative guidelines, and pilot projects focused on deepfake-resilient brand safety frameworks.

Theoretical Contributions

This bibliometric analysis theoretically organizes a fragmented research topic at the interface of brand safety, deepfakes, and synthetic media. The keywords, authors, and institutional networks indicate that the majority of deepfake research is primarily situated within cybersecurity and deep learning, or within broader discussions of AI, commerce, and the metaverse, with a limited yet expanding focus on explicitly linking these matters to brand-level outcomes such as trust, credibility, engagement, and loyalty. The study illustrates that brand safety in the deepfake era is better comprehended as a multifaceted ecosystem issue, encompassing technical detection capabilities, platform governance, legal frameworks, and consumer interpretation. The results endorse the creation of integrative models that connect (a) technological affordances and vulnerabilities (e.g., deep learning, detection), (b) platform and ecosystem frameworks (e.g., social media, metaverse), and (c) branding elements (e.g., authenticity, brand animosity, brand safety, crisis management). This study redefines brand safety from a limited "ad adjacency" framework to a more comprehensive theory of synthetic-media risk in branding, providing a structured array of research avenues for systematic future theory development.

Limitations

This study, like all bibliometric research, has several limitations that must be recognized. The outcomes are contingent upon the selection of the database and the search methodology (keywords, temporal parameters, document categories). Work published in non-indexed venues, publications, or practitioner reports is excluded, and research employing varied terminology for analogous phenomena may be inadequately represented. The analysis primarily examines English-language papers, potentially skewing the map in favor of Anglophone and Indian academic communities while underrepresenting contributions from locations where deepfake and brand safety issues are pertinent but published in other languages. Third, VOSviewer-generated maps exhibit sensitivity to parameter configurations (e.g., minimum citation thresholds, co-occurrence cri teria); varying thresholds may produce slightly divergent clusters, and the interpretation of these clusters is inherently subjective. The study is descriptive rather than evaluative; it does not evaluate the quality of individual studies or conduct a comprehensive content or thematic analysis of cons tructs and methodologies. Subsequent research may rectify these limitations by broadening the corpus across several databases and languages, integrating bibliometrics with systematic literature reviews or qualitative coding, and empirically validating the integrative conceptual frameworks proposed by the networks in this study.

CONCLUSION

This bibliometric analysis aimed to delineate the evolving knowledge landscape about brand safety in the deepfake era, integrating disparate research on cybersecurity, deep learning, synthetic media, and marketing. The findings indicate that the literature is structured around two primary focal points. A dense applied cluster connects deepfake technology, cybersecurity, artificial intelligence, commerce, and the metaverse, suggesting that researchers are progressively contextualizing deepfakes within AI-driven commercial and platform ecosystems. A distinct technical cluster focuses on deep learning and "cutting edges," highlighting swift methodological advancements in detection algorithms that remain inadequately linked with branding and consumer-oriented discussions. The collaborative maps substantiate this trend. Co-authorship and

affiliation networks indicate a limited number of closely connected teams that integrate information systems, marketing, and "digital futures" sectors, implying the emergence of interdisciplinary groups equipped to tackle deepfake threats from both technological and managerial perspectives. India stands forth as a significant nexus between Canada, the United Kingdom, and Gulf nations, but other areas seem more marginal. This geographical concentration signifies the locus of contemporary conceptual and empirical agenda-setting, while simultaneously underscoring the necessity for various viewpoints from alternative markets and regulatory environments. The findings indicate that brand safety in the deepfake age should be understood as a multifaceted ecosystem issue encompassing detection technologies, platform governance, regulatory frameworks, and consumer reactions. The maps highlight the necessity for practitioners to develop cross-functional competencies that integrate cybersecurity and AI skills with brand management and communication. For scholars, they indicate distinct deficiencies: insufficient theoretical integration between technical and branding literatures; a scarcity of empirical studies on consumer and stakeholder responses to deepfake brand incidents; and inadequately examined contexts, including non-Western markets, metaverse platforms, and influencer ecosystems. Bridging these gaps necessitates enhanced collaboration across disciplines and nations, alongside mixed-method designs that go from descriptive mapping to explanatory and predictive models of synthetic-media risk in branding.

REFERENCES

[1] B. TARCZYDŁO, "SOCIAL MARKETING PROJECTS IN BRAND ACTIVITIES. CASE STUDY.," *Sci. Pap. Silesian Univ. Technol. Organ. Manag. Nauk. Politech. Sl. Ser. Organ. i Zarz.*, no. 230, 2025.

[2] A. Diamantopoulos *et al.*, "Cost-effectiveness of an insertable cardiac monitor to detect atrial fibrillation in patients with cryptogenic stroke," *Int. J. Stroke*, vol. 11, no. 3, pp. 302–312, 2016.

[3] G. L. Bakris, "The role of combination antihypertensive therapy and the progression of renal disease hypertension: looking toward the next millennium," *Am. J. Hypertens.*, vol. 11, no. S7, pp. 158S-162S, 1998.

[4] L. Grewal, A. Stephen, and P. Vana, "Brands in unsafe places: effects of brand safety incidents on brand outcomes," *J. Mark. Res.*, 2025.

[5] C. L. Cofino, "Risks in the Digital Age," *J. homepage www. ijrpr. com ISSN*, vol. 2582, p. 7421.

[6] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. L. Rev.*, vol. 107, p. 1753, 2019.

[7] J. A. Goldstein and A. Lohn, "Deepfakes, Elections, and Shrinking the Liar's Dividend," *Brennan Cent. Justice, January*, vol. 23, 2024.

[8] K. J. Schiff, D. S. Schiff, and N. S. Bueno, "The liar's dividend: can politicians claim misinformation to evade accountability?," *Am. Polit. Sci. Rev.*, vol. 119, no. 1, pp. 71–90, 2025.

[9] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021.

[10] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, no. 4, pp. 959–975, 2017.

[11] N. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.

[12] I. Amerini and R. Caldelli, "Exploiting prediction error inconsistencies through LSTM-based classifiers to detect deepfake videos," in *Proceedings of the 2020 ACM workshop on information hiding and multimedia security*, 2020, pp. 97–102.

[13] R. Saxena and E. Gayathri, "Cyber threat intelligence challenges: Leveraging blockchain intelligence with possible solution," *Mater. Today Proc.*, vol. 51, pp. 682–689, 2022.

[14] M. Liu, J. Wang, X. Qian, and H. Li, "Audio-visual temporal forgery detection using embedding-level fusion and multi-dimensional contrastive loss," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 6937–6948, 2023.

[15] A. Jellali, I. Ben Fredj, and K. Ouni, "Data Augmentation for Convolutional Neural Network DeepFake Image Detection," in *2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, IEEE, 2023, pp. 1–5.

[16] I. Sharma, K. Jain, A. Behl, A. Baabdullah, M. Giannakis, and Y. Dwivedi, "Examining the motivations of sharing political deepfake videos: the role of political brand hate and moral consciousness," *Internet Res.*, vol. 33, no. 5, pp. 1727–1749, 2023.

[17] M. A. Salam, S. M. N. Rayun, W. Islam, R. Hasan, E. A. Firmansyah, and K. Kalinaki, "Consumer engagement: exploring deepfake applications in consumer marketing communication," in *Navigating the World of Deepfake Technology*, IGI Global, 2024, pp. 397–421.

[18] J.-Y. Shin, J. Suk, and J.-E. Chung, "Consumer Responses to fashion in the metaverse: A text-mining analysis on online

news comments," in *Future of Information and Communication Conference*, Springer, 2023, pp. 12–24.

[19] F. Sardana, K. K. Mishra, A. Singh, and N. Saini, "Transforming social media marketing through deepfake technology," in *Navigating the World of Deepfake Technology*, IGI Global, 2024, pp. 431–453.