Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.258270
Published by: Zeus Press

Research Evolution and Hotspots Analysis of Disinformation Induced by Generative Artificial Intelligence (2020–2025)

Hancheng Zheng*

School of Economics & Management, Nanjing Tech University, Nanjing 210000, China

**Corresponding author: Hancheng Zheng.*

Abstract

The ubiquitous integration of social media in our day to day life has one inevitable consequence, namely the serious threat of artificial intelligence (AI)-produced fake news and misinformation. Generative models can nowadays generate natural looking text with high quality. They are used in many different settings, and the consequences affect everyone which from individual users up through organizations, institutions, and societies. Monitoring where that research is going is not just an academic exercise, which requires identifying real hotspots. The data set for this analysis consists of 571 papers in English published by the Web of Science Core Collection, exclusively in 2020-2025 period. Multidimensional quantitative data of the literature analysis are acquired by CiteSpace and VOSviewer. It can be seen that there is a significant increase in publication after 2023. Currently, the United States and China are dominant in world production. Oddly enough, while overall macro-level productivity is high, institutions barely collaborate with each other at the micro-level. The thematic focus has shifted. The earlier works were largely chasing computer vision tasks such as detecting deepfakes; today's literature is heavy on controlling multimodal disinformation generated by LLMs. This trend is already seeping into risky domains, especially health care and journalism. Even with the emergence of new trends, fundamental knowledge here still resides on generative adversarial networks, residual networks, and preliminary deepfake evaluations. Finally, our study provides insight into the landscape of what is known within the discipline. Policy makers and scientists have a data-driven reference point from which to identify gaps in research and design policies that truly integrate science.

Keywords

generative artificial intelligence, disinformation, deepfake, large language models, bibliometrics, network analysis

1. Introduction

Generative AI is transforming the way in which content is produced. Existing models produce text/media that is practically indistinguishable from human [1], posing a complicated set of regulatory and ethical issues; researchers have cautioned against weaponizing them to mislead with fake data or smear reputations as early as 2018 [2]. Engineered disinformation is no longer merely distorting public perception—it divides society further,

even incites violence in some instances: fuels geopolitical strife [3]. With disinformation harder and harder to detect, growing mistrust in news organizations has made this issue one of high international concern [4].

This type of material is usually detected in one of the following ways: Content-based approaches are based on natural language processing as well as deep learning models such as RNNs, we deploy CNNs or Transformers for identifying syntactic oddities or sentiment anomalies [5]. The other solution is augmented by resources based checking, which performs spot checks of particular assertions with respect to other databases/knowledge bases [6].

Important technical challenges persist, though. Naıve binary classifiers fail routinely against the richness and diversity of manipulation schemes observed in the wild. Scaling is a further significant limitation, as slow inference times can prevent deployment of the model over huge social networks [7]. In short, LLMs still hallucinate: they frequently spit out reasonable sounding yet entirely false information – a deadly problem when applied in areas such as health care and law [8]. Human cognitive bias makes this dynamic even riskier, given that people tend to trust a confidently wrong AI more than a system that explicitly admits uncertainty. This creates an ethical trap where model optimization might prioritize user satisfaction over actual truth [9]. Even with mitigations like feature fusion or retrieval-augmented generation (RAG), unknown boundaries on what the LLMs do or don't detect [10].

Most existing work relies on one approach—either fine-tuning an algorithm, or debating within a narrow scope of ethics—which leaves a gap for understanding the larger research ecology of generative AI disinformation: Where are we doing this research? What kinds of underlying knowledge structures underlie it, and where is the frontier topic going in the future?

To fill that gap, we use bibliometrics in a quantitative mapping of the literature about AI for disinformation: Where is the core work published? In which country or institution? Identify impactful journals, and track the topic development of a field. Stepping away from individual technical surveys, we construct an overall picture of the literature, while using software such as CiteSpace or VOSviewer allows us to grasp the bigpicture overview which is absent in individual articles.

From a quantitative perspective, we can see that the two countries with the largest number of studies are the USA and China; the universities or institutes with the most significant contribution are Harvard Medical School and Zhejiang University, while the main publishers are also Journal of Medical Internet Research and IEEE Access. As for concept clusters, It is clear that the three major pillars are: deepfake detection, AI ethics, and health information. Moreover, from keyword analysis we can see there was a drastic change over time. Between 2020 and 2022, computer vision issues dominated the conversation. By the 2023–2025 window, the focus had pivoted heavily toward LLM-driven multimodal threats and their intersection with applied disciplines like healthcare.

This work sets up an actionable map to help guide further research in this area, as it highlights major authors and journals which drive the field. Tracing back conceptual clusters and keywords through time, the work illustrates how concerns about fake videos evolve into nuanced and context-dependent questions that need regulation for today's multimodal systems; it can provide a reference point for future design efforts in this area.

2. Methodology

2.1 Data Source

We limited our search in the literature exclusively to the Web of Science (WoS) Core Collection from the SCI-E and SSCI databases. WoS is an internationally known index which offers us access to all the data - references, authorship, funding, and citations—required to accurately map scientometrics [11]. To capture the intersection between AI and fake news, we iteratively merged the synonyms from each field together. The expanded keyword strings that were tuned in an initial search for maximum recall are presented in Table 1.

Table 1: List of the key words and the corresponding sets.

Set	Words Categories	Search Words
1	Basic concepts	Generative Artificial Intelligence, Generative AI, AIGC, Large Language Models
2	Functions	ChatGPT, Text Generation, Image Generation, Multimodal Generation
3	Risks	False information, Fabricated information, Fake news, Misinformation, Misleading

		information, Deepfakes, Illusions, Digital fraud

2.2 Data Acquisition

We searched titles, abstracts, and author keywords using the WoS advanced search. Our general query captured different terminology to describe AI deception albeit with many false positives, that required manual screening. Our first 2015–2025 search returned 1,340 English articles and reviews. Reading the titles and abstracts uncovered one major confounding group of papers: those about building AI to detect old-school human disinformation, not necessarily focusing on disinformation generated by artificial intelligence, but we discarded those as well as nonpeer-reviewed content such as editorials and conference proceedings, leaving us with a final workable corpus of 571 peer-reviewed articles and reviews published between 2020 and 2025.

2.3 Analysis Tools and Methods

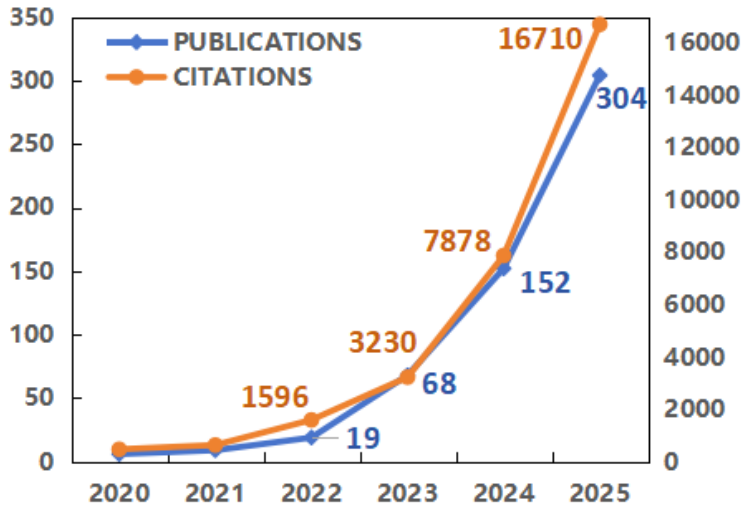
To decode the structure and evolution change, we integrated two scient metric instruments – CiteSpace and VOSviewer. The main reason why we choose to utilize CiteSpace was that it can be considered as a longitudinal tracking instrument with functions including burst detection and betweenness centrality calculation, thus being able to identify the milestone paper(s) or turning points on the citation map; we applied this method into co-citation maps, co-word networks, and time zone diagrams; the co-occurrence maps were also generated in VOSviewer. Although CiteSpace deals with dynamic evolution, the visualization of dense clusters and structure closeness are better performed by VOSviewer. Combined, they complement each other's analytic blind spots.

3. Specific Analysis Results and Discussion

3.1 Analysis of Annual Publication Output

Figure 1 presents the number of published papers, as well as citations, in a given year by all the selected publications (N=571), between years 2020-2025. First articles about AI based fake news have been presented during early 2020, followed by a small but steady growth of the literature up to and including 2022; however, an important inflection point took place during 2023, with an exponential increase of publications. The growth reached its peak in year 2025 with the publication of 304 articles resulting in 16,710 references. Together these metrics point to a clear indication that deepfake misinformation research is quickly becoming an important and influential field in the wider body of work, a trend that will probably continue for some time.

Figure 1: Distribution of publications and citations in AI-Generated misinformation



3.2 Disciplinary Distribution and Co-occurrence

Figure 2 & Figure 3 show major fields that respond to AI fake news, as well as the association network of each field. From quantitative results it can be seen clearly that there is an emphasis on technological areas. CS, Information Systems has contributed the most articles (n=128) followed closely by Engineering, Electrical & Electronic with n=125). While Computer Science, Artificial Intelligence adds another 76 entries. Disciplines in this cluster represent core technology areas that are also concerned about issues surrounding machine-generated content (e.g., false information, faulty results). These disciplines operate within well-defined boundaries where verification systems for falsehoods and faulty outcomes can play a role.

Figure 2: Distribution of the major subjects in AI-Generated misinformation

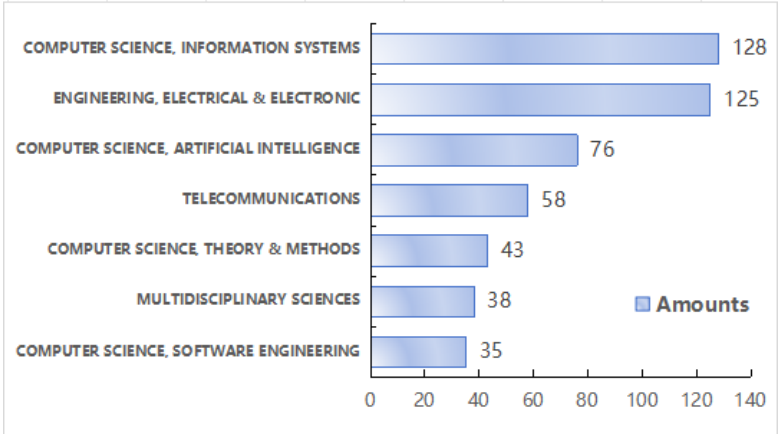
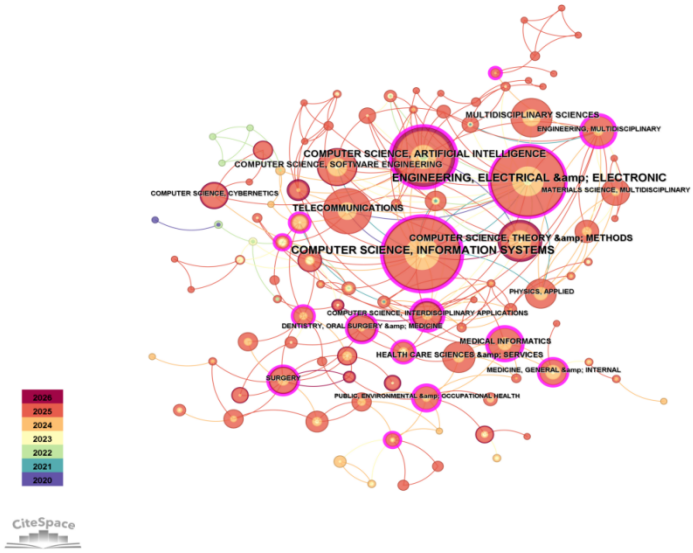


Figure 3: Main subject network in AI-Generated misinformation

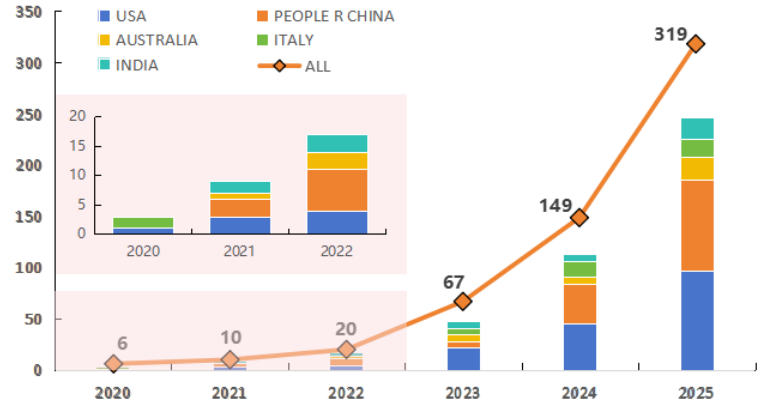


3.3 High-Yield Countries and Global Collaboration Networks

International collaboration analysis reflects the evolution trend in spreading of knowledge, as well as where is the geographical focus of this area of study. Fig. 4 shows a map of the number of publications for all active countries between 2020-2025. Five most productive nations - USA, China, Australia, Italy, India - these four countries produce almost half (48.49 %) of the world's production. The US and China stand out here: a clear edge on the rest for publication output, and

We divide previous works about AI-based misinformation according to time periods. The first period is from 2020-2022 and during this stage, this area was just at an early stage; publication numbers were rather small and constant, mainly dealing with setting up theory; whereas the third period (2023–2025) is characterised by a sharp increase of publications. This exponential growth shows an abrupt change and makes this subject become a very active research area due to new technology demands.

Figure 4: Distribution of publications of major producing countries



The top-ten countries are described in Table 2. High values of citation per paper, for some countries, indicate that papers from these countries have a high citation rate.

Figure 5: Country cooperation network in AI-Generated misinformation

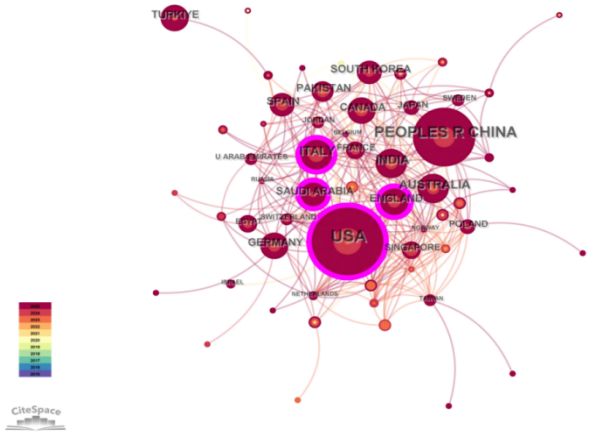


Table 2: Top 10 most productive countries

Rank	Country	TP	TC	TC/TP	Frequency	Centrality	Year
1	**USA**	175	8702	49.73	168	0.16	2020-2025
2	China	130	7740	59.54	132	0.02	2021-2025
3	Australia	35	3113	88.94	40	0.04	2021-2025
4	**Italy**	41	2758	67.27	39	0.11	2020-2025
5	India	35	2888	82.51	38	0.06	2021-2025
6	**England**	36	2208	61.33	31	0.14	2020-2025
7	Germany	30	1991	66.37	29	0.1	2023-2025
8	South Korea	32	1881	58.78	29	0.04	2021-2025
9	Saudi Arabia	25	1897	75.88	29	0	2021-2025
10	Canada	27	1643	60.85	27	0.13	2022-2025

Note: TP = total number of publications; TC = total citations of the TP; Frequency= frequency of the node in the network; Centrality = betweenness centrality of the node in the network; Year = time span from the first paper to the last paper. Bold font indicates countries that are more central and have made greater contributions to the development of the international collaboration network.

3.4 Core Research Institutions and Academic Hubs

By mapping out their output and collaborations, we identify major driving forces in this field from an academic perspective. There are a total of 208 nodes and 253 links within the institutional co-authorship graph related to generative AI–induced disinformation (Figure 6). Generally speaking, the most central institutions with respect to both number of publications and betweenness tend to be the structural leaders. Particularly,

Harvard Medical School (frequency=10, centrality=0.14), Zhejiang University (10, 0.01) hub in the network, initiating many international cooperative projects. For example, Tianjin University (7, 0.05) and the University of Sydney (6,0.05), and they are important institutions too.

Notably, the maximum date of publication for this elite set falls very near to that time: during 2022-2025. That's also aligned with the rise and spread of LLMs, showing how quickly the research community was able to respond to these new ethical and security concerns that arose in consequence of this event.

Figure 6: Institutional cooperation network in AI-Generated Misinformation

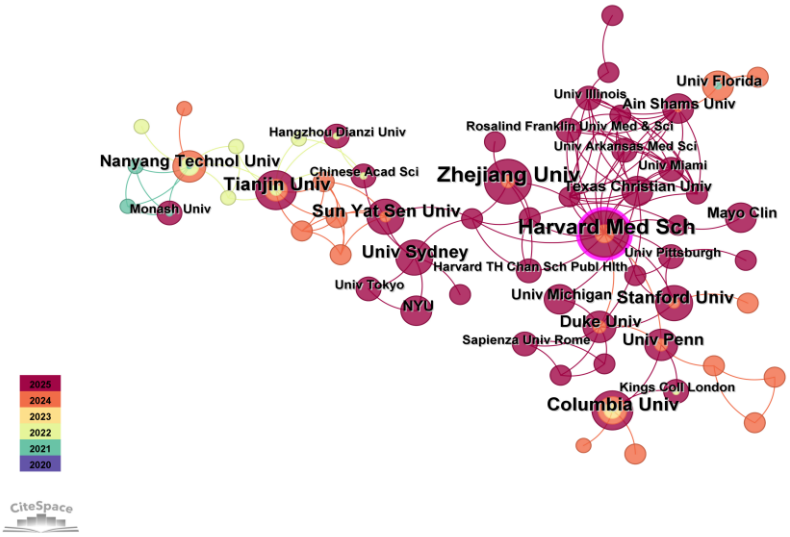


Table 3: Top 10 most prolific institutions.

Rank	Institution	Country	TP	TC	TC/TP	Frequency	Centrality	Year
1	Harvard Med Sch	USA	11	481	43.73	10	0.14	2023-2025
2	Zhejiang University	China	10	650	65.00	10	0.01	2024-2025
3	Tianjin Univ	China	11	856	77.82	7	0.05	2022-2025
4	Columbia Univ	USA	8	275	34.38	7	0.01	2022-2025
5	Univ Sydney	Australia	8	436	54.50	6	0.05	2023-2025
6	Stanford Univ	USA	7	231	33.00	6	0.01	2023-2025
7	Duke Univ	USA	7	313	44.71	5	0.03	2024-2025
8	Nanyang Technol Univ	Singapore	6	634	105.67	5	0.03	2021-2025

3.5 Core Author and Collaboration Networks

Figure 8 is used to visualize the collaboration pattern and the topological structure of some key authors, there are obvious local clusterings but no overall center in this network; rather functioning as sets of disjointed sub-communities (i.e., the Javed–Irtaza, Elhusseiny– Brown, and Li–Gao clusters). Similar to Table 3, these stand-alone pieces indicate that present production occurs only within the confines of institutional groups and there are clear impediments at a higher level for international collaboration. In addition, most of node-creation and link-creation takes place in 2023–2025, an abrupt growth that coincides with the surge of large generative artificial intelligence (AI) models lately.

The most influential of these are the top ten core authors listed Table 4. Javed has the highest overall output and citations (5, 284) followed by his coauthor Irtaza (4,240). On the other hand, the mean effect is maximum for Guarnera with a value of 70.00 references/article than that of Irtaza with 60.00 and Javed with 56.80 which shows their effective work in this area from the technological point of view.

Figure 8: Author cooperation network in AI-Generated Misinformation

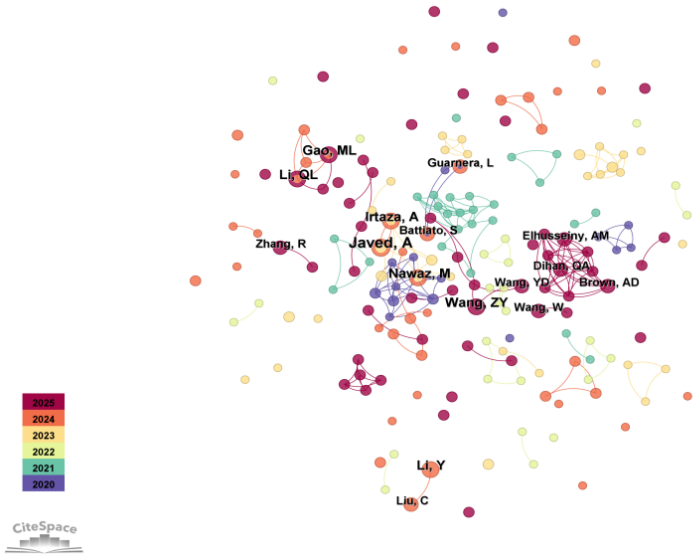


Table 4: Top 10 most influential authors.

Rank	Author	Country	Affiliation	TP	TC	TC/TP	Frequency	Centrality	Year
1	Javed, Ali	Pakistan	UET Taxila	5	284	56.80	5	0.00	2023-2024
2	Elhusseiny, Abdelrahman M.	USA	Univ Arkansas Med Sci	4	138	34.50	3	0.00	2024-2025
3	Li, Qilei	China	Cent China Normal Univ	4	168	42.00	4	0.00	2024-2025
4	Zhang, Guisheng	China	Shandong Univ Technol	4	168	42.00	2	0.00	2024-2025
5	Irtaza, Aun	Pakistan	UET Taxila	4	240	60.00	4	0.00	2023-2024
6	Gao, Mingliang	China	Shandong Univ Technol	4	168	42.00	4	0.00	2024-2025
7	Brown, Andrew D.	USA	Univ Arkansas Med Sci	4	138	34.50	3	0.00	2024-2025
8	Hatipoglu, Sirin	Unknown	Istanbul Univ Cerrahpasa	3	119	39.67	2	0.00	2025
9	Guarnera, Luca	Italy	Univ Catania	3	210	70.00	3	0.00	2020-2024
10	Danesh, Arman	Canada	Western Univ	3	35	11.67	1	0.00	2023-2025

3.6 Analysis of Highly Cited Core Journals

The mainstream of AI-CV community is high-quality papers in AI/CV fields. The discussions of the AIGC-powered disinformation rely on ML advances, pattern recognition, neural networks. At the same time, we rely heavily on the arXiv preprint archive to obtain new attack-defense algorithms or LLM evaluation datasets; it also indicates this field is fast-iterating and timely. Finally, the threat posed by generative AI disinformation is no longer an issue for niche scientific communities alone but one with wide-reaching, multidisciplinary science involvement.

3.7 Analysis of Highly Cited Core Papers

The top ten most bursting cited articles are listed in Table 5 and can be used to trace shifts in research focus over time within the domain. The article by Goodfellow et al. [12] on GANs is one such article. Its maximum burst strength was 6.38, focuses most heavily in 2020-2022, where researchers focused on core mechanics of generative algorithms as a first step toward creating defenses against disinformation. We observe clear patterns in citations. Papers cite classical GANs and CV basics heavily in 2020–2022; with deepening of deepfakes from 2022 to 2024, the emphasis moved. The spikes in citations from that second wave emphasize structural detection techniques [13], as well as general evaluations of more recent AIGC attacks such as Deepfakes [14]. This is in line with a general shift in other research areas to abandon single-task models for more sophisticated multi-task models.

We can see that the document co-citation network has a very dense value, and also has a strong recency bias. Both suggest the immediacy of interest on such researches. In addition, the structure of this network is strongly affected by technology ancestry. Although the present review comes from 2020-2025, the most important nodes indeed refer to those in first generation of Deepfake (from 2018 to 2020). The backbone of this knowledge bank

are the papers that introduced first and/or large forgery datasets; or major key generation techniques. We can see that, in the network view, those older predecessors keep their large degrees of structural importance, surrounded by a lot of recent citing papers on one side until years 2024–2025.These consistent hit rates suggest an obvious interpretation, namely that contemporary attempts at detecting massive-model AIGC disinformation continue to rely upon those initial benchmark corpora and original adversarial models. In spite of the quick deployment of these new generation generators, these first contributions are still very relevant nowadays.

Figure 9: Network of most influenced journals

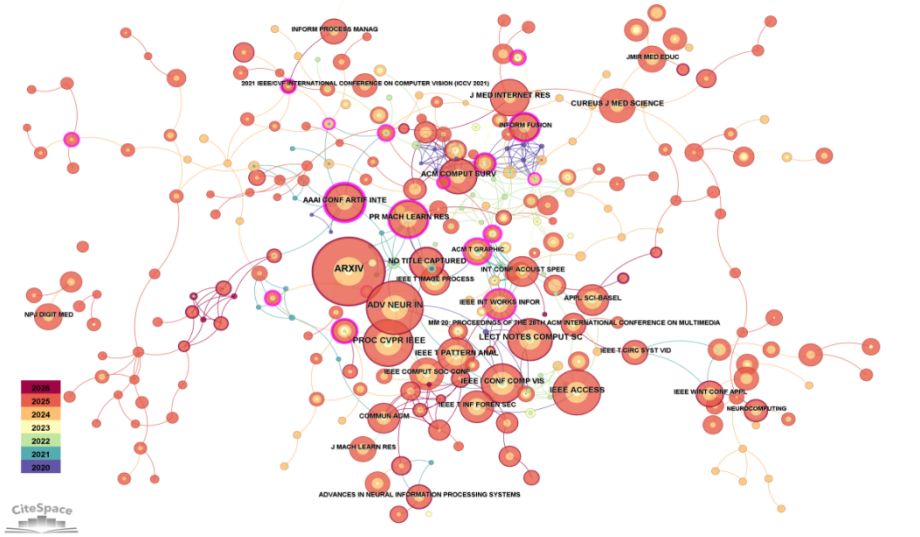


Figure 10: Network of most cited papers

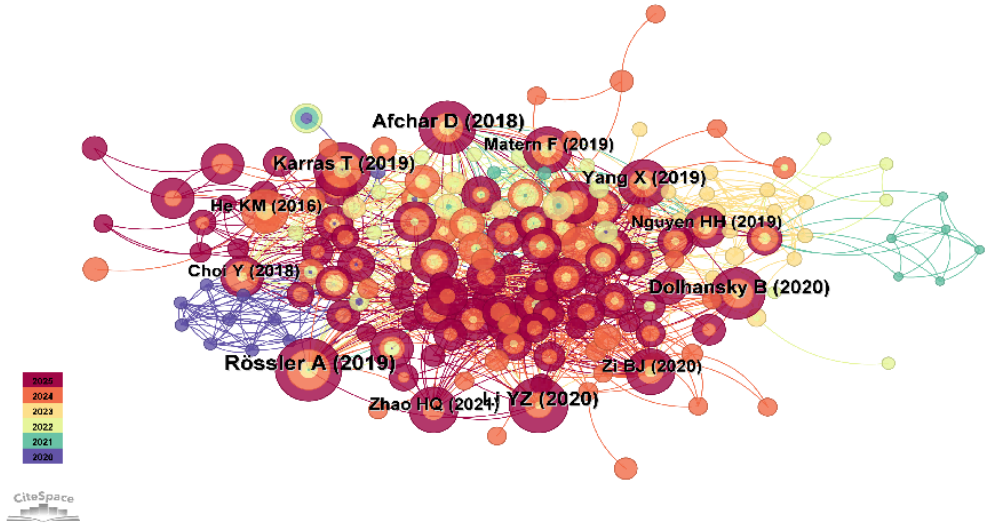


Table 5: Top 10 most cited papers.

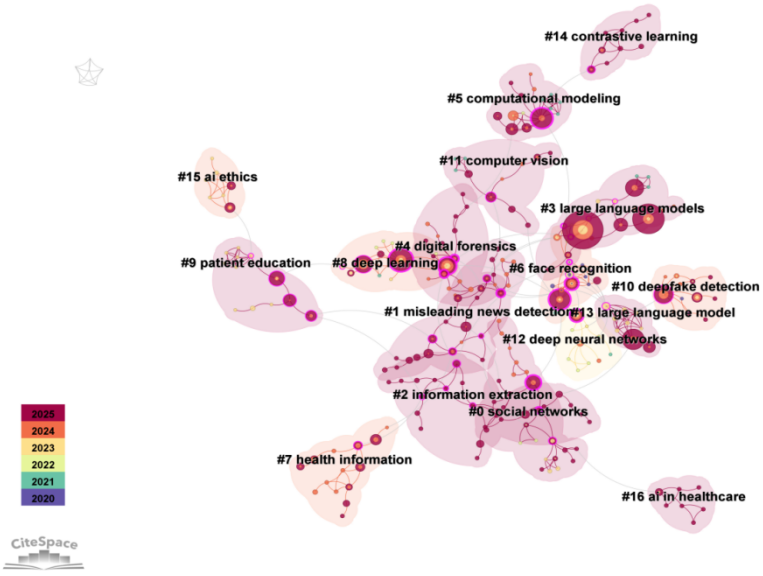
References	Strength	Begin	End	2015-2025
Goodfellow IJ(2014) [15]	6.38	2020	2022	
Verdoliva L(2020) [16]	3.02	2020	2022	
Kemelmacher-Shlizerman I(2017) [17]	5.14	2021	2023	
Hsu CC(2020) [18]	4.76	2021	2022	
Zhang Ying(2017) [19]	4.2	2021	2023	
Li YZ(2018) [20]	3.27	2021	2022	
He KM(2016) [21]	4.83	2022	2024	
Liu ZW(2015) [22]	2.91	2022	2023	
Neves JC(2020) [23]	2.91	2022	2023	
Masood M(2023) [24]	4.07	2023	2024	

4. Research Hotspots and Frontier Analysis

4.1 Reference Co-citation Network

To describe the background knowledge stock and their evolution process, we carried out the reference co-citation analysis. To ensure both visual intuitiveness and effective description, we adopted an adaptive 8-year window length to calculate all the references of each cited document as well. As shown in, the resulting graph has 459 vertices (weighted according to cocitation counts) and 620 weighted time edges; its topology is well supported, with the presence of a large value of modularity Q = 0.a high number of clusters formed (c = 8616) as well as a high weighted average silhouette value (S=0.9488), indicating good internal cluster cohesion.

Figure 11: Reference co-citation network clusters in AI-Generated misinformation



We conducted a clustering of the co-citation graph to identify patterns on which ways knowledge in this domain developed, and it returned 17 large clusters, were labeled from #0 to #16 according to their total volume. We annotated them using LLR applied to reference titles and WoS keywords. Observing the network structure of Fig. 11, the architecture is relatively dense, and the dominant warm color indicates an obvious burst of paper in the period of 2023-2025 Large Model era.

The 17 clusters broadly agree to four categories. The cluster 0, 3 and 13 locate exactly at the centre of the network's mass; its central location reflects a strong influence by LLM hallucination defects on macro-level fake news contents which is more persuasive than that written by human [25]. Technical-side clusters (Clusters 8, 10, and 11) also overlap heavily. This overlap represents the current efforts in developing reference databases [6] and designing more sophisticated forensics techniques including contrastive learning [7, 8], spatiotemporal feature extraction [9].to combat multimodal deepfakes [26]. Beyond simple algorithms, two clusters emerge separately: Cluster 7 (Health Information), and Cluster 9 (Patient Education). This second group points out that health care is one of the most critical areas in which unsafe mistakes from AI systems could put people's lives at risk [27]. Finally, Cluster 15 (AI Ethics), sit at the edge but are nonetheless foundational; reflecting a major trend in the field: There is more interest to move away from pure technical defence towards socio-technical governance and safety alignment.

4.2 Keyword Co-occurrence Analysis and Temporal Evolution

In order to visualize how such a scientific field has evolved over time as well as its structural changes we generated a dynamic word association map using a timezone chart (Fig. 12). The size of each node represents the number of keywords: while edge distribution maps inter-annual conceptual linkages.

The experimental results cover an important turning point: from loose connectedness in 2020, to clustered high-density connectivity in 2025. Before 2023, the paradigm was completely grounded on computer vision-

based counter-measures. The common jargon at that time included “face manipulation”, “generative adversarial networks” (GAN), etc., which, incidentally, happens to be a vanilla tech-defense which was localized to the specific problem of identity-theft/facial-synthesis [28].

This local focus disappeared after 2023. The abrupt emergence of “large language model”, “diffusion models” indicates the structural shift towards multimodal disinformation – in particular, textto-video pipelines [29]. Notably, we see that the expansion vector has now reached the real world biosecurity realm with a periphery node “health” and “cardiac arrest.” From CS to pandemic: Fabricated algorithms now undermine medical advic [30] the need for interdisciplinarity and timely action.

Figure 12: Keyword co-occurrence network in AI-Generated misinformation

5. Conclusion

This evaluation maps the evolution of AI-generated disinformation research using 571 WoS records (2020–2025). A core transition occurred post-2023: research expanded from basic visual manipulation to the multimodal complexities introduced by LLMs and diffusion models. Consequently, technical defences in practice: the use of such techniques to reduce actual risk in critical domains, e.g., public health or journalism.

From a structural point of view, we can also find some problems in this research field. In terms of number of papers published, it is dominated by the USA and China; however, their individual betweenness centralities are both zero. It has many small clusters. Scholars work independently from different universities or institutes, without substantive international cooperation. And to make things worse, our theoretical basis is built upon outdated networks, poses a theoretical bottleneck to evaluating modern LLMs.

The findings on both health data and responsible use of AI indicate there are not silver bullet solutions for dealing with those challenges and tackling such problems require tighter coupling of technology design and policy making together. The significance of enhancing approaches (e.g., CL and RAG) indicates improved multilingual models would benefit fact-checkers. Importantly, developing accountability and data provenance standards in dangerous domains will rely upon proactive, interdisciplinary activity across computer scientists, attorneys; physicians; and lawyers.

References

[1] Dwivedi, Y. k., Kshetri, N., Hughes, L., Slade, E. louise, Jeyaraj, A., Kar, A. kumar, Baabdullah, A. m., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. ahmad, Al-Busaidi, A. s., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of

generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*(102642). https://doi.org/10.1016/j.ijinfomgt.2023.102642

[2] Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Comput. Surv.*, *54*(1), 7:1-7:41. https://doi.org/10.1145/3425780

[3] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023a). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

[4] Zou, Q.-Y., Chen, G., Wu, X.-K., Huang, P., & Chen, M. (2025). AIE-FND: Enhancing UAV-Based Fake News Detection Technique via AI-Generated Insights From Media Experts. *Ieee Transactions on Consumer Electronics*, *71*(4), 10965–10976. https://doi.org/10.1109/TCE.2025.3603976

[5] Zhou, F., Yan, X., Wang, Y., & Xiong, Y. (2025). Collaborative Governance and Decision Support Strategies for Misinformation Generated by Generative Artificial Intelligence. *Journal of Organizational and End User Computing*, *37*(1), 1–24. https://doi.org/10.4018/JOEUC.397671

[6] Zou, Q.-Y., Chen, G., Wu, X.-K., Huang, P., & Chen, M. (2025). AIE-FND: Enhancing UAV-Based Fake News Detection Technique via AI-Generated Insights From Media Experts. *Ieee Transactions on Consumer Electronics*, *71*(4), 10965–10976. https://doi.org/10.1109/TCE.2025.3603976

[7] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023b). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

[8] Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *630*(8017), 625–630. https://doi.org/10.1038/s41586-024-07421-0

[9] Bar-Or Nirman, D., Weizman, A., & Azaria, A. (2025). Fool Me, Fool Me: User Attitudes Toward LLM Falsehoods. *Ieee Access*, *13*, 200351–200366. https://doi.org/10.1109/ACCESS.2025.3632748

[10] Ma, Y., Zhang, X., Ren, J., Wang, R., Wang, M., & Chen, Y. (2025). Linguistic features of AI mis/disinformation and the detection limits of LLMs. *Nature Communications*, *17*(1). https://doi.org/10.1038/s41467-025-67145-1

[11] Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

[12] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[14] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023a). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

[15] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

[16] Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE journal of selected topics in signal processing*, *14*(5), 910–932.

[17] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). *FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces* (arXiv:1803.09179). arXiv. https://doi.org/10.48550/arXiv.1803.09179

[18] Hsu, C.-C., Zhuang, Y.-X., & Lee, C.-Y. (2020). Deep fake image detection based on pairwise learning. *Applied Sciences*, *10*(1), 370.

[19] Zhang, Y., Zheng, L., & Thing, V. L. (2017). Automated face swapping and its detection. *2017 IEEE 2nd international conference on signal and image processing (ICSIP)*, 15–19. https://ieeexplore.ieee.org/abstract/document/8124497/

[20] Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*. http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Li_Exposing_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.pdf

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[22] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of the IEEE international conference on computer vision*, 3730–3738. http://openaccess.thecvf.com/content_iccv_2015/html/Liu_Deep_Learning_Face_ICCV_2015_paper.html

[23] Ciftci, U. A., Demir, I., & Yin, L. (2020). How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals. *2020 IEEE international joint conference on biometrics (IJCB)*, 1–10. https://ieeexplore.ieee.org/abstract/document/9304909/

[24] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023b). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

[25] Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, *9*(26), eadh1850. https://doi.org/10.1126/sciadv.adh1850

[26] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE/CVF international conference on computer vision*, 1–11. http://openaccess.thecvf.com/content_ICCV_2019/html/Rossler_FaceForensics_Learning_to_Detect_Manipulated_Facial_Images_ICCV_2019_paper.html

[27] Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887. https://www.mdpi.com/2227-9032/11/6/887

[28] Crothers, E. N., Japkowicz, N., & Viktor, H. L. (2023). Machine-generated text: A comprehensive survey of threat models and detection methods. *IEEE Access*, *11*, 70977–71002.

[29] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

[30] Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, *47*(1), 33. https://doi.org/10.1007/s10916-023-01925-4

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.