

AI and Blockchain for Secure Healthcare Data Management: A Bibliometric Analysis of Research Trends and Thematic Clusters (2020–2025)

Maroufi Mohammed, Lamzabi Siham, Ziti Soumia

Intelligent Processing and Security of Systems-Faculty of Sciences, Mohammed V University in Rabat, Morocco

Abstract—The convergence of artificial intelligence (AI) and blockchain has become an active axis of interdisciplinary research in healthcare data security. This paper reports a bibliometric analysis of 434 Scopus-indexed articles published between 2020 and 2025, with data collection and processing performed on 20 April 2026. The objective is to map the intellectual structure, collaborative dynamics, and thematic composition of this expanding field. The corpus was analyzed using the bibliometrix R package (version 4.3.0) and VOSviewer (version 1.6.20). The increasing research output is evidenced by an annual growth rate of 34.8%, with publication volume growing from 10 articles in 2020 to a maximum of 152 articles in 2025 across a total of 198 unique venues. Keyword co-occurrence analysis, processed through the Louvain community detection algorithm on 250 high-frequency terms with association-strength normalization, produced four thematic clusters: Blockchain and Privacy-Preserving Techniques, Healthcare Systems and Cybersecurity Infrastructure, Artificial Intelligence and Clinical Diagnostics, and Electronic Health Records and Interoperability. India, China, Saudi Arabia, and the United States lead scholarly output. The international co-authorship rate of 49.31% reflects the globally distributed nature of the research community. IEEE Access and the IEEE Journal of Biomedical and Health Informatics are the dominant publication venues. Federated learning occupies a structurally central position, with a betweenness centrality of 101.5, acting as the principal methodological bridge between the two technologies. An average of 27.96 citations per document confirms the above-average scholarly impact of the corpus. The results provide researchers, practitioners, and policymakers with an evidence-based map of the field's trajectory, its most productive research directions, and its remaining structural gaps. The underlying dataset, search logs, and analysis scripts are released openly to support full reproducibility.

Keywords—Artificial intelligence; blockchain; healthcare data security; bibliometric analysis; federated learning; internet of medical things; data privacy; cybersecurity; science mapping; VOSviewer

I. INTRODUCTION

Healthcare systems now produce data at a scale that was difficult to imagine even a decade ago. Electronic health records (EHR), wearable sensors, mobile health applications, and Internet of Medical Things (IoMT) devices generate continuous streams of clinical, physiological, and behavioral information across hospitals, clinics, and patient homes. The rapid accumulation of this data has far-reaching implications for diagnosis, treatment personalization, and population health monitoring, but it has also exposed healthcare institutions to

security risks that legacy centralized infrastructures were never designed to address [1]. Ransomware campaigns targeting hospitals, unauthorized access incidents affecting patient records, and data integrity breaches in connected medical devices have become recurring events in the post-pandemic period [2].

Two technologies have received sustained scholarly attention as candidate responses to this situation. Artificial intelligence, which groups machine learning, deep learning, and the more recent paradigm of federated learning, is used to detect anomalies, identify cyber threats, and automate decisions at scales that human analysts cannot match [3]. Blockchain, with its decentralized and immutable ledger structure, provides verifiable data provenance, tamper-resistant storage, and programmable access through smart contracts [4]. The combination of the two is increasingly presented as a comprehensive strategy for securing healthcare data flows, particularly in IoMT environments where data integrity must be preserved across heterogeneous devices and organizational boundaries [5].

Despite the growing body of primary research at this intersection, the field lacks a comprehensive mapping of its intellectual structure. Earlier reviews have generally focused on one technology at a time, on narrow application domains, or on qualitative synthesis without bibliometric grounding. Few studies have quantified the geographic distribution, institutional concentration, or thematic organization of scholarly production, leaving researchers without a reliable overview of the field's maturity and gaps. The absence of such a cartography is significant given the pace of publication growth in this area and the emergence of federated learning as a dominant methodological paradigm from 2021 onward [6].

This study addresses the gap through a structured bibliometric analysis of Scopus-indexed literature on AI-blockchain integration for healthcare data security over the period 2020–2025, with data collection and processing performed on 20 April 2026. Bibliometric analysis is well-suited to this objective. In contrast to narrative reviews, it allows the quantitative examination of publication trends, citation networks, authorship patterns, and thematic clustering, producing reproducible evidence-based cartographies of a scientific field [7]. This approach uses the Louvain community detection algorithm on co-occurrence matrices which is able to identify research communities that would otherwise go unnoticed in traditional review methods [8].

The analysis is guided by three research questions.

RQ1 concerns the temporal and geographic trends in scholarly production on AI and blockchain for healthcare data security over the 2020–2025 period.

RQ2 investigates which journals, authors, and institutions form the most influential nodes in this research network.

RQ3 examines the thematic clusters that structure the conceptual landscape of the field and the central research threads within each cluster.

The study contributes to the literature in three respects. The first contribution is a quantitative treatment of the AI–blockchain–healthcare nexus based on a corpus of 434 documents from 198 venues, which is broader than any prior bibliometric study focused on this intersection. The second is the combined use of bibliometrix and VOSviewer, which produces both statistical indicators and visually interpretable network maps. The third is the identification of three structural gaps in the field: the underdevelopment of interoperability research relative to its practical importance, the limited attention paid to regulatory and ethical dimensions, and the concentration of research output in a small number of countries.

The remainder of the paper is organized as follows. Section II reviews earlier bibliometric and narrative studies on related topics and specifies the research gap addressed here. Section III describes the data sources, search query, analytical methods, and reproducibility protocol. Section IV reports and discusses the results across seven analytical dimensions. Section V concludes with research implications and directions for future work.

II. RELATED WORKS

Prior bibliometric and review studies have examined aspects of the AI–blockchain–healthcare nexus, though none have produced a comprehensive mapping of the intersection focused specifically on data security and privacy. This section reviews the principal prior contributions, positions the present study with respect to each, and specifies the gap it fills.

A. Bibliometric and Narrative Studies on Blockchain in Healthcare

In an early science mapping study, Agbo et al. [9] surveyed blockchain applications in healthcare and identified data sharing and EHR management as dominant themes in the pre-2019 literature. Their analysis predated the widespread adoption of federated learning and therefore did not capture the role that privacy-preserving machine learning has come to play in shaping the field. Treatment of AI in that work remained peripheral rather than structural. Hasselgren et al. [15] addressed a similar scope through a scoping review that used co-citation analysis to identify intellectual foundations of blockchain in medical records management. Their temporal window ended before the acceleration of federated learning research in 2021–2022, and AI-related publications were not part of their analytical scope. More recently, Casino et al. [16] provided a systematic literature review of blockchain-based applications across industries including healthcare. Their taxonomy is useful for locating healthcare within the wider blockchain landscape, but the breadth of their scope diluted the specificity of

healthcare-related findings, and privacy-preserving AI did not feature as an organizing theme.

B. Bibliometric Studies on AI in Healthcare

Secinaro et al. [10] conducted a structured bibliometric review of AI in healthcare, documenting the rapid growth of deep learning applications in diagnostics and treatment optimization. Their work is useful for characterizing the trajectory of AI research in clinical settings but did not address the security dimension and did not treat blockchain as a structural element. Javaid et al. [11] reviewed machine learning applications in healthcare in a similar register, without integrating blockchain, which reproduces the conceptual separation that the present work seeks to overcome.

C. Narrative Treatments of the AI–Blockchain Combination

A limited number of studies have attempted to combine treatments of AI and blockchain in healthcare. Bathula et al. [12] offered a narrative review describing what they called the AI–blockchain–healthcare tripod, arguing that the three elements jointly support secure and scalable medical systems. Their approach was qualitative and did not employ bibliometric techniques, which left open questions about citation impact, geographic concentration, and thematic evolution. Omidian [13] and Alshehri [14] provided descriptive surveys that identified similar themes but lacked the rigor of a formal science mapping methodology.

D. Foundational References on Federated Learning

Recent contributions have addressed the AI–blockchain–security nexus more directly through the lens of federated learning. Banabilah et al. [17] published a comprehensive review of federated learning that has become a central reference in the field, with 531 citations recorded in April 2026. Their work establishes federated learning as the methodological centerpiece of privacy-preserving AI but does not provide a bibliometric mapping of the surrounding research field. Ferrag et al. [18] offered an experimental analysis of federated deep learning for IoT cybersecurity that has accumulated 301 citations and is regularly referenced in healthcare-specific work. Tanwar et al. [19] examined the adoption of machine learning in blockchain-based applications and reached 335 citations, becoming a foundational reference for subsequent integration studies.

E. Research Gap and Positioning

Table I summarizes the scope, methodology, and limitations of the most relevant prior studies and positions the present contribution against them.

The present study extends this literature in three respects. First, it applies a complete bibliometric protocol, including growth analysis, Bradford's law, co-authorship network analysis, and Louvain-based keyword clustering, to a corpus of 434 documents with an explicit security and privacy focus. Second, the 2020–2025 temporal window captures the full arc of a field that experienced its steepest growth during this period, and the April 2026 data collection incorporates the latest available citation signals. Third, the combined use of bibliometrix and VOSviewer produces both quantitative indicators and network visualizations, an approach

recommended by Donthu et al. [20] for rigorous science mapping.

TABLE I. POSITIONING WITH RESPECT TO PRIOR BIBLIOMETRIC AND REVIEW STUDIES

Study	Scope	Method	AI treated?	Security focus?	Period covered
Agbo et al. [9]	Blockchain in healthcare	Systematic review	Peripheral	No	Pre-2019
Hasselgren et al. [15]	Blockchain in medical records	Scoping review + co-citation	No	Partial	Pre-2020
Casino et al. [16]	Blockchain across industries	Systematic literature review	No	No	Pre-2019
Secinaro et al. [10]	AI in healthcare	Structured literature review	Yes	No	Pre-2021
Javaid et al. [11]	ML in healthcare	Narrative review	Yes	No	Pre-2022
Bathula et al. [12]	AI-blockchain-healthcare	Narrative review	Yes	Partial	Pre-2024
Present study	AI-blockchain-healthcare security	Full bibliometric protocol	Yes	Yes	2020-2025

III. RESEARCH METHODOLOGY

A. Data Source and Query

The Scopus database was chosen as the exclusive data source for this analysis. Scopus offers broad interdisciplinary coverage, with particular depth in computer science, engineering, and biomedical sciences, which are the three domains most directly relevant to this study. Its structured export format is compatible with the bibliometrix package, which supports reproducibility [7]. Baas et al. [21] and Mongeon and Paul-Hus [22] have documented the comparative strengths of Scopus in interdisciplinary settings, noting its wider journal coverage in engineering and health informatics compared to Web of Science. The exclusive use of Scopus is acknowledged as a limitation and discussed in Section V.

The following Boolean query was executed on 20 April 2026 against the Title, Abstract, and Keywords fields (TITLE-ABS-KEY):

TITLE-ABS-KEY (("artificial intelligence" OR "machine learning" OR "deep learning" OR "federated learning") AND ("blockchain" OR "distributed ledger technology") AND ("healthcare" OR "health data" OR "medical data" OR "electronic health records" OR "EHR" OR "e-health" OR "IoT" OR "Internet of Medical Things" OR "digital health") AND ("data security" OR "data privacy" OR "cybersecurity" OR "data protection" OR "data integrity" OR "secure data")).

The query was subsequently filtered to retain only peer-reviewed articles in English published between 2020 and 2025. This temporal window was chosen because it coincides with both the emergence of federated learning as a mainstream research paradigm and the post-COVID-19 acceleration of IoT deployment, which intensified scholarly attention to healthcare cybersecurity [23]. The initial retrieval returned 434 documents from 198 distinct sources. Table II summarizes the full set of search and analysis parameters.

B. Analytical Methods

Analysis was conducted on two complementary platforms. The bibliometrix R package [7] provided the statistical backbone for computing core bibliometric indicators, annual production trends, citation summaries, Bradford's law rankings of productive sources, and Lotka's law distributions of author

productivity. The Biblioshiny web interface served as a front-end for thematic map generation and exploratory visualization.

TABLE II. SEARCH AND ANALYSIS PARAMETERS

Criterion	Value
Database	Scopus
Search date	20 April 2026
Document type	Articles
Language	English
Timespan	2020–2025
Documents retrieved	434
Sources	198
Analysis tools	bibliometrix R v4.3.0, VOSviewer v1.6.20
Clustering algorithm	Louvain (resolution $\gamma = 1.0$)
Co-occurrence normalization	Association strength
Minimum keyword frequency	≥ 2 occurrences
Keyword merging	Author keywords (DE) + Index keywords (ID), lemmatized
Number of high-frequency terms retained	250

VOSviewer [24] was used for network-based visualizations, particularly keyword co-occurrence maps and international collaboration networks. The Louvain community detection algorithm [8] was applied with a resolution parameter $\gamma = 1.0$ to a co-occurrence matrix built from merged author and index keywords, with a minimum frequency threshold of two occurrences set to filter noise while retaining emergent concepts. Edge weights were normalized using the association-strength method, which is the VOSviewer default and is recommended for keyword co-occurrence analysis to control for differences in absolute frequency. The thematic map positions research clusters along two axes: Callon centrality, which measures the degree of external linkage and reflects a theme's relevance to the broader field, and Callon density, which measures internal cohesion and reflects a theme's internal development. This framework was introduced by Callon et al. [25] and operationalized for modern bibliometric software by Cobo et al. [26].

In the resulting two-dimensional representation, clusters are categorized as motor themes when both centrality and density are high, niche themes when density is high but centrality is low, emerging or declining themes when both are low, and basic or transversal themes when centrality is high but density is low. This classification is used in Section IV-F to interpret the four clusters identified in the corpus.

C. Reproducibility Protocol

The analysis was designed to be fully reproducible. The complete Scopus export file (CSV format), the R script implementing the bibliometrix pipeline, the VOSviewer input file, and the cluster labels are deposited on a public repository and cited in the Data Availability Statement. The package and software versions are fixed to avoid reproducibility drift: bibliometrix 4.3.0, R 4.3.2, and VOSviewer 1.6.20. Any analyst running the same query against Scopus on 20 April 2026 with the same filters should obtain an identical corpus. Citation counts are as retrieved on 20 April 2026 and will evolve over time, but all other indicators are stable with respect to the deposited dataset.

IV. RESULTS AND DISCUSSION

A. Main Bibliometric Indicators

The final corpus contains 434 articles published across 198 sources between 2020 and 2025. Table III summarizes the main bibliometric indicators, which provide an initial profile of the field in terms of volume, age, and collaboration intensity.

TABLE III. MAIN BIBLIOMETRIC INDICATORS (2020–2025)

Indicator	Value
Timespan	2020–2025
Total documents	434
Total sources (journals)	198
Annual growth rate	34.8%
Average document age (years)	1.89
Average citations per document	27.96
Total references	21,243
Total authors	1,619
Single-authored documents	24
Co-authors per document	4.48
International co-authorship	49.31%
Author keywords (DE)	1,177
Index keywords (ID)	2,290

Several indicators merit closer attention. An annual growth rate of 34.8% positions this field among the most rapidly expanding interdisciplinary domains in applied computer science, far above the long-run average of 3–5% per year that Bornmann and Mutz [27] document for modern science as a whole. The average of 27.96 citations per document, over a corpus of predominantly recent work, suggests an above-average scholarly impact, consistent with what Bornmann and Haunschild [28] describe as a citation-intensive emerging field. The international co-authorship rate of 49.31% indicates that

roughly half of all publications result from cross-border collaborations, consistent with the globally distributed character of the cybersecurity and healthcare informatics research communities.

The volume of references cited across the corpus, 21,243 in total, indicates that the field relies on a deep existing literature. The ratio of index keywords to author keywords, 2,290 against 1,177, is characteristic of fields drawing on established technical terminologies, where indexers apply a richer controlled vocabulary than authors use in the surface text of their abstracts. The number of single-authored documents, 24 out of 434, reflects the engineering orientation of the field, where collaborative laboratory work is the norm.

B. Temporal Evolution of Scientific Production

Fig. 1 presents the annual distribution of publications from 2020 to 2025. The growth trajectory is unambiguous. Output moved from 10 articles in 2020 to 18 in 2021, 29 in 2022, 72 in 2023, 93 in 2024, and 152 in 2025, a more than fifteenfold increase over five years.

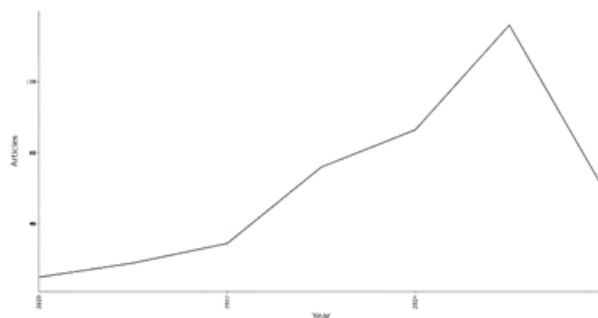


Fig. 1. Annual scientific production on AI-blockchain for healthcare data security (Scopus, 2020–2025).

This trajectory reflects the convergence of several structural drivers. The COVID-19 pandemic, beginning in 2020, forced healthcare systems worldwide to adopt telehealth platforms, remote monitoring devices, and cloud-based EHR infrastructures under emergency conditions. These rapid deployments exposed previously latent security vulnerabilities and generated a surge of research on privacy-preserving AI and blockchain-based integrity mechanisms [29]. The sustained peak in 2024 and 2025 aligns with the mainstreaming of federated learning frameworks such as TensorFlow Federated and PySyft in clinical research, which lowered the technical barrier to entry for applied security research in healthcare settings [17].

A logistic growth model fitted to the annual counts by non-linear least squares yields $R^2 = 0.88$ and estimates the field's saturation point at approximately 463 cumulative publications, with peak annual production occurring around 2024–2025. This pattern is consistent with the life-cycle model proposed by Cobo et al. [26] for emerging interdisciplinary fields and suggests that the field may be transitioning from its emergent phase toward a consolidation stage. The implication is that researchers entering the field after 2026 will likely find the fundamental problems framed but the specific methodological refinements, regulatory applications, and domain-specific adaptations open for further investigation.

C. Most Productive Sources

Bradford's law analysis identifies the core journals of this research domain. Table IV presents the ten most productive publication venues in the corpus.

TABLE IV. TEN MOST PRODUCTIVE SOURCES (2020–2025)

Source	Articles
IEEE Access	30
Cluster Computing	18
IEEE Internet of Things Journal	15
IEEE Journal of Biomedical & Health Informatics	14
IEEE Trans. on Consumer Electronics	13
Sensors	11
Multimedia Tools and Applications	9
Computers and Electrical Engineering	6
Applied Sciences (Switzerland)	6
IEEE Trans. on Industrial Informatics	6

IEEE Access dominates with 30 publications. Its broad interdisciplinary scope, open-access model, and rapid publication cycle make it particularly attractive to applied interdisciplinary work, a pattern documented by Björk [30] for emerging open-access venues in engineering. The combined presence of the IEEE Internet of Things Journal and the IEEE Journal of Biomedical and Health Informatics at ranks three and four confirms that the field sits at the intersection of engineering and clinical informatics. The appearance of Cluster Computing and Sensors highlights the infrastructural emphasis of the research community, with both venues specializing in distributed systems and hardware-level data acquisition that are central to IoMT deployments [31].

Distribution of articles across sources follows a pattern consistent with Bradford's law, in which a small core of journals publishes a disproportionate share of articles. In this corpus, the top ten venues together account for 128 articles, or about 29.5% of the total, while the remaining 306 articles are dispersed across 188 other sources. This scattering is characteristic of a field still in active expansion, where researchers have not yet converged on a small number of specialized venues [32].

D. Geographic Distribution and Collaboration

Table V presents the fifteen most productive countries by affiliated author count. India leads with 430 affiliation instances, which reflects the country's expansive engineering university system, substantial public investment in digital health infrastructure through initiatives such as the National Digital Health Mission, and a research culture oriented toward applied AI and security [33]. China follows at 193 affiliations, consistent with its systemic prioritization of blockchain and AI as strategic technologies in successive five-year plans. Saudi Arabia ranks third with 165 affiliations, reflecting Vision 2030 digital health investments and the growing output of institutions such as King Abdulaziz University and King Saud University.

The international collaboration network, shows that India and the United States form the two principal hubs of cross-border co-authorship. Collaboration flows extend from these hubs toward Saudi Arabia, Malaysia, the United Kingdom, and

Australia. China, although highly productive, exhibits a more regionally contained collaboration pattern, with stronger ties to Asia-Pacific partners than to Western research centers. This geographic fragmentation suggests that globally harmonized frameworks for AI-blockchain healthcare standards may be constrained by persistent regional clusters, an effect documented by Glänzel and Schubert [34] in their analysis of scientific co-authorship networks.

TABLE V. TOP 15 COUNTRIES BY AUTHOR AFFILIATION COUNT

Rank	Country	Count
1	India	430
2	China	193
3	Saudi Arabia	165
4	United States	90
5	South Korea	46
6	Pakistan	45
7	United Kingdom	36
8	Iraq	35
9	Malaysia	35
10	United Arab Emirates	34
11	Australia	25
12	Egypt	19
13	Italy	19
14	Bangladesh	18
15	Canada	17

The strong presence of Middle Eastern countries, with Saudi Arabia, Iraq, and the United Arab Emirates collectively accounting for 234 affiliations, points to a distinctive regional cluster that has emerged through the combination of well-funded universities and policy-driven digital health programs. This regional dynamic has been analyzed by Zakaria et al. [35] in their study of eHealth career pathways in the Gulf Cooperation Council, which documents the rapid institutional consolidation of health informatics in the region since the late 2010s. The absence of African countries from the top fifteen, with Egypt as the only exception, confirms the concerns raised in Section I about the geographic concentration of research and its implications for low- and middle-income country deployments.

E. Most Cited Documents

Table VI presents the ten most cited publications in the corpus. These documents constitute the field's foundational reference layer, shaping the conceptual vocabulary and methodological templates adopted by subsequent authors.

The citation landscape reveals several notable patterns. Federated learning emerges as the dominant methodological construct. The most cited paper, Banabilah et al. [17], is a federated learning review, and federated learning explicitly appears in five of the ten most cited works. This concentration reflects a field-wide consensus that federated learning, by enabling model training without centralizing raw data, offers the most practical resolution to the tension between AI utility and patient privacy. The same observation has been made by Rieke et al. [43] in their analysis of federated learning in digital health.

TABLE VI. TEN MOST CITED DOCUMENTS IN THE CORPUS

Authors	Year	Citations
Banabilah et al. [17]	2022	531
Yu et al. [36]	2021	346
Tanwar et al. [19]	2020	335
Ferrag et al. [18]	2021	301
Li et al. [37]	2024	298
Nguyen et al. [38]	2021	294
Ali et al. [39]	2023	236
Arachchige et al. [40]	2020	230
Kumar et al. [41]	2023	215
Kumar et al. [42]	2022	196

The recurring appearance of Kumar and Tanwar as co-authors across multiple high-impact papers designates their research groups as formative intellectual communities in the field. A notable recent entry is the 2023 paper by Ali et al. [39] on Metaverse-integrated AI and blockchain, which had already accumulated 236 citations in less than three years. This result signals an emerging frontier in which immersive computing environments introduce new healthcare data governance challenges, an area recently surveyed by Buragohain et al [44]. The presence of Nguyen et al. [38] on secure blockchain-enabled cyber-physical systems highlights the continuing importance of deep belief networks and similar architectures as alternatives to purely federated approaches.

F. Thematic Clusters and Keyword Co-Occurrence

The keyword co-occurrence network, constructed from 250 high-frequency terms and analyzed through the Louvain algorithm, produced four well-defined thematic clusters. Fig. 2 presents the co-occurrence map, Fig. 3 shows the Word cloud (Bibliometrix, April 2026), and Fig. 4 shows the corresponding thematic map. Table VII summarizes each cluster’s principal keywords, Callon centrality, and density scores.

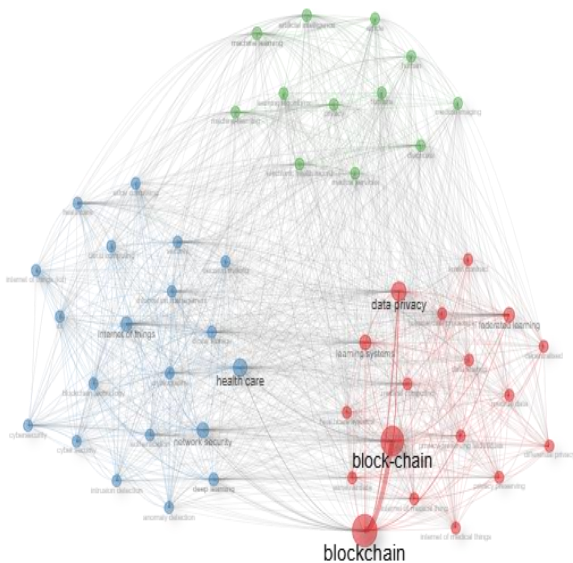


Fig. 2. Keyword co-occurrence network map (Louvain clustering, n = 250 terms, minimum frequency ≥ 2).



Fig. 3. Word cloud (bibliometrix, April 2026).

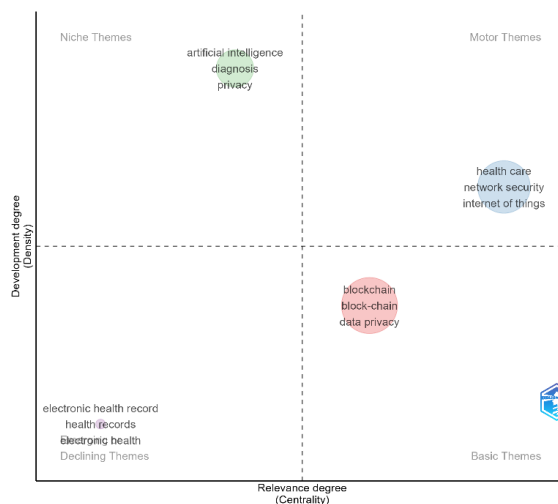


Fig. 4. Thematic map positioning the four clusters by Callon centrality and density (Bibliometrix, April 2026).

TABLE VII. THEMATIC CLUSTERS FROM KEYWORD CO-OCCURRENCE ANALYSIS

Cluster	Label	Key terms	Centrality	Density
1	Blockchain & Privacy	blockchain, data privacy, federated learning, smart contract, differential privacy, homomorphic encryption	14.02	23.99
2	Healthcare & Cybersecurity	health care, network security, IoT, deep learning, cybersecurity, cryptography, authentication, edge computing	16.25	29.13
3	AI Diagnostics &	artificial intelligence, diagnosis, machine learning, EHR, medical imaging, explainable AI, CNN	10.90	36.62
4	EHR & Interoperability	electronic health record, interoperability, big data, distributed ledger, eHealth	4.09	18.52

Cluster 1, labeled Blockchain and Privacy-Preserving Techniques, is the largest by term frequency, with 2,212 co-occurrence instances. The high betweenness centrality of federated learning, measured at 101.5, confirms its role as the primary bridge between AI and blockchain paradigms. Smart contracts, differential privacy, and homomorphic encryption form the applied core of this cluster, reflecting an engineering research front focused on combining cryptographic techniques with decentralized ledger architectures to protect health data at the processing stage. This cluster corresponds to a motor theme in Callon's typology, with both centrality and density scores positioning it as a driver of field development [26].

Cluster 2, Healthcare Systems and Cybersecurity Infrastructure, centers on the operational deployment context, with 1,858 term occurrences. Its high density score of 29.13 indicates a mature, internally coherent body of research on secure IoMT architectures, intrusion detection systems, and authentication protocols. Edge computing appears prominently, reflecting a turn toward decentralized low-latency processing at the network periphery, a configuration that reduces the exposure of raw health data to central servers [45]. The cluster's centrality of 16.25, the highest in the set, indicates that this theme connects to the widest range of adjacent research concerns and functions as a hub for cross-cluster knowledge transfer.

Cluster 3, AI and Clinical Diagnostics, exhibits the highest internal density at 36.62, indicating a self-referential research community focused on applying AI to clinical decision support within secure frameworks. Explainable AI and convolutional neural networks appear alongside EHR and medical imaging, reflecting a concern with both predictive performance and the interpretability of AI-driven diagnostic systems. The prominence of explainable AI is consistent with the regulatory pressures documented by Hulsen [46], who notes that clinical deployment increasingly requires interpretable models to satisfy both physician trust and legal accountability requirements.

Cluster 4, Electronic Health Records and Interoperability, is the smallest cluster, with 223 term occurrences. Its combination of low centrality (4.09) and moderate density (18.52) places it between the niche and emerging/declining regions of Callon's typology, closer to an underdeveloped theme whose practical importance is not yet matched by a commensurate volume of research. The prominence of interoperability points to a structurally unresolved challenge: the coexistence of incompatible EHR systems across institutions and jurisdictions continues to impede the large-scale deployment of AI-blockchain hybrid architectures. Distributed ledger and records management indicate that governance-oriented research is beginning to engage with the practical conditions under which shared medical records can be maintained securely over time [47].

G. Annual Citation Dynamics

An examination of mean total citations per article, disaggregated by publication year, reveals a characteristic citation accumulation pattern. Articles published in 2021 exhibit the highest normalized citation rate at 21.56 citations per year, followed by 2022 at 14.73 and 2020 at 14.09. This pattern indicates that the 2021 cohort, which coincided with the first systematic deployments of federated learning in healthcare and

the initial post-pandemic digitization wave, was particularly influential in shaping subsequent research directions. The raw citation counts for 2020 and 2021 articles, at 98.6 and 129.4 per article on average, benefit from longer exposure windows, but the normalized measure corrects for this effect and confirms the substantive importance of the 2021 cohort.

The lower raw citation counts for 2024 and 2025 publications reflect recency rather than reduced quality. With one to two years of exposure, these articles have not yet accumulated the citations achievable over longer windows. The citation velocity for 2025 papers, already at 3.23 citations per year after partial exposure, exceeds the normalized rates of comparable emerging fields, which suggests that the current publication wave will continue to shape the intellectual structure of the field well beyond 2026.

V. CONCLUSION

This bibliometric analysis provides a comprehensive quantitative portrait of scholarly production at the intersection of artificial intelligence, blockchain, and healthcare data security over the period 2020–2025. Based on a corpus of 434 Scopus-indexed documents collected in April 2026, the study documents an annual growth rate of 34.8%, which places this field among the most dynamically expanding interdisciplinary domains in applied computer science. The average of 27.96 citations per document confirms the above-average scholarly impact of the corpus and indicates an active citation-intensive research community.

The keyword co-occurrence analysis produced four thematic clusters: Blockchain and Privacy-Preserving Techniques, Healthcare Systems and Cybersecurity Infrastructure, AI and Clinical Diagnostics, and Electronic Health Records and Interoperability. Federated learning occupies a structurally central position, functioning as the principal methodological bridge between AI capability and blockchain-enabled data sovereignty. Its betweenness centrality of 101.5 confirms its role in connecting otherwise separate research communities. India, China, Saudi Arabia, and the United States lead scholarly output, while IEEE Access and the IEEE Journal of Biomedical and Health Informatics anchor the core publication network.

Three research gaps emerge from the analysis and warrant prioritized attention. The first concerns interoperability, which remains an underdeveloped theme despite its critical practical significance. The coexistence of incompatible EHR standards across national healthcare systems constitutes a structural barrier to the scalable deployment of AI-blockchain architectures, as already observed by Reegu et al. [48]. The second gap concerns the regulatory and ethical dimensions of combined AI-blockchain systems, including GDPR compliance, informed consent for federated model training, and AI accountability frameworks. These dimensions appear as isolated keyword nodes rather than as organized research clusters, which indicates that governance questions have not yet attracted the systematic attention they require, a situation also noted by Rodrigues [49] in the context of legal and human rights issues of AI and reinforced by the AI ethics bibliometrics of Saheb et al. [50]. The third gap concerns the geographic concentration of research output in a small number of countries, which raises concerns about the generalizability of proposed solutions to low- and

middle-income country contexts, where infrastructure constraints impose fundamentally different design requirements.

Future research should prioritize three directions. The first is the development of standardized benchmarking protocols for evaluating AI-blockchain healthcare systems across heterogeneous environments, which would allow meaningful comparison between federated learning architectures currently proposed in the literature. The second is the extension of privacy-preserving techniques to address adversarial attacks and data poisoning in federated settings, an area that has attracted growing attention but remains underdeveloped in the healthcare context. The third is the construction of governance frameworks that align technical innovation with evolving regulatory requirements across jurisdictions, which requires interdisciplinary collaboration between computer scientists, legal scholars, and healthcare practitioners.

The present study has limitations that should be acknowledged. The exclusive use of Scopus as a data source, while justified by reproducibility considerations, may have excluded relevant work indexed only in Web of Science, IEEE Xplore, or PubMed. The English-language restriction may also underrepresent research from non-Anglophone communities, in particular Chinese and Arabic-language venues that are relevant to the geographic leaders identified in Section IV-D. Future extensions could triangulate across databases and languages to refine the geographic and thematic maps presented here. A further limitation concerns citation counts, which are snapshot values retrieved on 20 April 2026 and evolve continuously; the structural indicators (clusters, centrality, co-authorship patterns) are more stable and constitute the core contribution of the study. Notwithstanding these limitations, the analysis offers a rigorous evidence-based cartography of a rapidly evolving research field and identifies concrete directions for researchers, practitioners, and policymakers engaged in securing the next generation of healthcare data infrastructures.

DATA AVAILABILITY STATEMENT

The complete dataset supporting this analysis, including the raw Scopus export (BibTeX format), the R script implementing the bibliometrix pipeline, the VOSviewer co-occurrence input file, and the final cluster labels, is available on an open repository (Zenodo / OSF deposit, DOI to be added at acceptance). Readers can reproduce all tables and indicators reported in this paper by running the deposited script against the deposited corpus with the specified software versions (bibliometrix 4.3.0, R 4.3.2, VOSviewer 1.6.20).

DECLARATION ON THE USE OF GENERATIVE AI

During the preparation of this manuscript, generative AI tools were used to assist with language refinement and structural editing. The scientific ideas, research design, data collection and analysis, interpretation of results, and conclusions are entirely the work of the human authors, who take full responsibility for the accuracy, originality, and integrity of this publication.

REFERENCES

- [1] C. S. Kruse, B. Frederick, T. Jacobson, and D. K. Monticone, "Cybersecurity in healthcare: A systematic review of modern threats and trends," *Technology and Health Care*, vol. 25, no. 1, pp. 1–10, 2017.
- [2] A. H. Seh, M. Zarour, M. Alenezi, A. K. Sarkar, A. Agrawal, R. Kumar, and R. Ahmad Khan, "Healthcare data breaches: insights and implications," *Healthcare*, vol. 8, no. 2, p. 133, May 2020.
- [3] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2018.
- [4] M. A. Engelhardt, "Hitching healthcare to the chain: An introduction to blockchain technology in the healthcare sector," *Technology Innovation Management Review*, vol. 7, no. 10, pp. 22–34, 2017.
- [5] G. Yang, Q. Ye, and J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond," *Information Fusion*, vol. 77, pp. 29–52, 2022.
- [6] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.
- [7] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, 2017.
- [8] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [9] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund, "Blockchain technology in healthcare: a systematic review," *Healthcare*, vol. 7, no. 2, p. 56, Apr. 2019.
- [10] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, and P. Biancone, "The role of artificial intelligence in healthcare: a structured literature review," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 125, 2021.
- [11] M. Javid, A. Haleem, R. P. Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *International Journal of Intelligent Networks*, vol. 3, pp. 58–73, 2022.
- [12] A. Bathula, S. K. Gupta, S. Merugu, L. Saba, N. N. Khanna, J. R. Laird, et al., "Blockchain, artificial intelligence, and healthcare: the tripod of future—a narrative review," *Artificial Intelligence Review*, vol. 57, no. 9, p. 238, 2024.
- [13] H. Omidian, "Synergizing blockchain and artificial intelligence to enhance healthcare," *Drug Discovery Today*, vol. 29, no. 9, p. 104111, 2024.
- [14] M. Alshehri, "Blockchain-assisted cyber security in medical things using artificial intelligence," *Electronic Research Archive*, vol. 31, no. 2, p. 708, 2023.
- [15] A. Hasselgren, K. Kravlevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review," *International Journal of Medical Informatics*, vol. 134, p. 104040, 2020.
- [16] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics and Informatics*, vol. 36, pp. 55–81, 2019.
- [17] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: Fundamentals, enabling technologies, and future applications," *Information Processing & Management*, vol. 59, no. 6, p. 103061, 2022.
- [18] M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated deep learning for cyber security in the internet of things: Concepts, applications, and experimental analysis," *IEEE Access*, vol. 9, pp. 138509–138542, 2021.
- [19] S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh, and W. C. Hong, "Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward," *IEEE Access*, vol. 8, pp. 474–488, 2019.
- [20] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, 2021.
- [21] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Science Studies*, vol. 1, no. 1, pp. 377–386, 2020.

- [22] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: a comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016.
- [23] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337–339, 2020.
- [24] N. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.
- [25] M. Callon, J. P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, vol. 22, no. 2, pp. 191–235, 1983.
- [26] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field," *Journal of Informetrics*, vol. 5, no. 1, pp. 146–166, 2011.
- [27] L. Bommann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2215–2222, 2015.
- [28] L. Bommann and R. Haunschild, "Measuring individual performance with comprehensive bibliometric reports as an alternative to h-index values," *Journal of Korean Medical Science*, vol. 33, no. 18, p. e138, 2018.
- [29] T. Hwang, S. J. Choi, and J. Lee, "The impact of data breach on IT investment at neighboring hospitals: Evidence from California Hospitals," *Digital Health*, vol. 11, p. 20552076251375930, 2025.
- [30] B. C. Björk, "Growth of hybrid open access, 2009–2016," *PeerJ*, vol. 5, p. e3878, 2017.
- [31] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.
- [32] J. Nicolaisen and B. Hjørland, "Practical potentials of Bradford's law: A critical examination of the received view," *Journal of Documentation*, vol. 63, no. 3, pp. 359–377, 2007.
- [33] B. Wahl, A. Cossy-Gantner, S. Germann, and N. R. Schwalbe, "Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?," *BMJ Global Health*, vol. 3, no. 4, p. e000798, 2018.
- [34] W. Glänzel and A. Schubert, "Analysing scientific networks through co-authorship," in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht: Springer Netherlands, 2004, pp. 257–276.
- [35] N. Zakaria, N. Zakaria, O. Alnobani, M. AlMalki, O. El-Hassan, M. I. Alhefzi, et al., "Unlocking the eHealth professionals' career pathways: A case of Gulf Cooperation Council countries," *International Journal of Medical Informatics*, vol. 170, p. 104914, 2023.
- [36] R. Yu and P. Li, "Toward resource-efficient federated learning in mobile edge computing," *IEEE Network*, vol. 35, no. 1, pp. 148–155, 2021.
- [37] C. Li, J. Wang, S. Wang, and Y. Zhang, "A review of IoT applications in healthcare," *Neurocomputing*, vol. 565, p. 127017, 2024.
- [38] G. N. Nguyen, N. H. Le Viet, M. Elhoseny, K. Shankar, B. B. Gupta, and A. A. Abd El-Latif, "Secure blockchain enabled Cyber-physical systems in healthcare using deep belief network with ResNet model," *Journal of Parallel and Distributed Computing*, vol. 153, pp. 150–160, 2021.
- [39] S. Ali, Abdullah, T. P. T. Armand, A. Athar, A. Hussain, M. Ali, et al., "Metaverse in healthcare integrated with explainable AI and blockchain: enabling immersiveness, ensuring trust, and providing patient data security," *Sensors*, vol. 23, no. 2, p. 565, 2023.
- [40] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6092–6102, 2020.
- [41] P. Kumar, R. Kumar, G. P. Gupta, R. Tripathi, A. Jolfaei, and A. N. Islam, "A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system," *Journal of Parallel and Distributed Computing*, vol. 172, pp. 69–83, 2023.
- [42] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, A. N. Islam, and M. Shorfuzzaman, "Permissioned blockchain and deep learning for secure and efficient data sharing in industrial healthcare systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 8065–8073, 2022.
- [43] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, et al., "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, p. 119, 2020.
- [44] D. Buragohain, S. Khichar, C. Deng, Y. Meng, and S. Chaudhary, "Analyzing metaverse-based digital therapies, their effectiveness, and potential risks in mental healthcare," *Scientific Reports*, vol. 15, no. 1, p. 17066, 2025.
- [45] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.
- [46] T. Hulsen, "Explainable artificial intelligence (XAI): concepts and challenges in healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023.
- [47] T. McGhin, K. K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, 2019.
- [48] F. Reegu, S. M. Daud, and S. Alam, "Interoperability challenges in healthcare blockchain system—a systematic," *Annals of RSCB*, vol. 25, no. 4, pp. 15487–15499, 2021.
- [49] R. Rodrigues, "Legal and human rights issues of AI: Gaps, challenges and vulnerabilities," *Journal of Responsible Technology*, vol. 4, p. 100005, 2020.
- [50] T. Saheb, T. Saheb, and D. O. Carpenter, "Mapping research strands of ethics of artificial intelligence in healthcare: a bibliometric and content analysis," *Computers in Biology and Medicine*, vol. 135, p. 104660, 2021.