

Federated Learning in Distributed Computing: A Scopus-Based Bibliometric Analysis of Research Trends

Loso Judijanto
IPOSS Jakarta

Article Info

Article history:

Received Apr, 2026

Revised Apr, 2026

Accepted Apr, 2026

Keywords:

Data Privacy;
Distributed Computing;
Edge Computing;
Federated Learning

ABSTRACT

This study presents a comprehensive bibliometric analysis of federated learning within distributed computing, aiming to explore research trends, intellectual structures, and emerging themes in the field. Data were retrieved from the Scopus database covering publications from 2010 to 2025. The analysis employs performance metrics and science mapping techniques, including co-authorship, keyword co-occurrence, citation, and density visualization using VOSviewer. The results reveal a significant increase in research output, particularly after 2018, driven by the growing demand for privacy-preserving machine learning and the expansion of edge computing and Internet of Things (IoT) ecosystems. Key research clusters focus on data privacy, system optimization, distributed machine learning, and real-world applications such as healthcare and industrial systems. The findings also highlight strong global collaboration networks and the dominance of contributions from leading countries such as China and the United States. Furthermore, recent trends indicate a shift toward integrating federated learning with advanced technologies such as blockchain, reinforcement learning, and energy-efficient systems. This study provides a structured overview of the field and offers valuable insights for researchers and practitioners in identifying research gaps and future directions in federated learning within distributed computing.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Name: Loso Judijanto

Institution: IPOSS Jakarta

Email: losojudijantobumn@gmail.com

1. INTRODUCTION

The rapid advancement of digital technologies has fundamentally reshaped how data is generated, processed, and utilized across diverse sectors. In the era of big data and pervasive connectivity, distributed computing has become a critical paradigm for managing large-scale and decentralized data environments. However, conventional centralized machine learning approaches—relying on aggregating data

into a single repository—are increasingly constrained by critical challenges, including privacy risks, security vulnerabilities, high communication overhead, and stringent regulatory requirements [1], [2]. These limitations are particularly evident in data-intensive domains such as healthcare analytics, financial technologies, smart cities, and Internet of Things (IoT) ecosystems, where data is inherently distributed and sensitive [3], [4]. Consequently, there is a pressing need for innovative computational

frameworks that enable collaborative intelligence while preserving data ownership and privacy.

Federated learning has emerged as a transformative solution to these challenges. Initially introduced by H. Brendan McMahan and colleagues at Google, federated learning enables multiple decentralized entities to collaboratively train machine learning models without sharing raw data. Instead, model updates such as parameters or gradients are exchanged, thereby minimizing data exposure and mitigating privacy risks. This decentralized paradigm aligns naturally with distributed computing infrastructures, where data resides across heterogeneous devices and networks [5], [6]. As a result, federated learning has gained significant traction in domains that demand both high analytical performance and strict compliance with data governance standards.

The growing relevance of federated learning is further reinforced by the increasing emphasis on privacy-preserving technologies and regulatory frameworks. Policies such as the General Data Protection Regulation (GDPR) have imposed strict obligations on organizations regarding data collection, storage, and processing [7]–[9]. These regulatory pressures have accelerated the adoption of decentralized learning paradigms, as organizations seek to extract value from data without violating legal and ethical constraints. Simultaneously, the rapid expansion of edge computing and IoT ecosystems has intensified the fragmentation of data sources, further necessitating scalable and efficient federated learning approaches.

Given the exponential growth of research in this domain, a systematic evaluation of the existing body of knowledge is essential. Bibliometric analysis has become a robust methodological approach for mapping scientific developments, identifying influential contributions, and uncovering emerging research trends. By utilizing large-scale academic databases such as Scopus, researchers can examine publication trajectories, citation structures, and collaboration networks. Furthermore, visualization tools like VOSviewer enable a

comprehensive exploration of relationships among authors, keywords, and institutions, thereby providing deeper insights into the intellectual structure and evolution of a research field.

Despite the increasing volume of studies on federated learning, there remains a notable gap in the literature regarding comprehensive bibliometric analyses that specifically explore its intersection with distributed computing. Existing studies predominantly emphasize technical advancements or domain-specific applications, often overlooking a holistic perspective that captures the intellectual landscape, thematic evolution, and global collaboration patterns. This limitation underscores the need for a systematic and data-driven investigation that elucidates the broader dynamics of federated learning research within distributed environments.

Therefore, this study aims to conduct a Scopus-based bibliometric analysis of federated learning in distributed computing. Specifically, this research seeks to: (1) analyze the temporal growth and distribution of publications, (2) identify key contributors, including authors, institutions, and countries, and (3) map dominant research themes and emerging trends through keyword co-occurrence and citation analysis. By offering a structured and comprehensive overview, this study contributes to advancing the understanding of federated learning's research trajectory and provides strategic directions for future scholarly and practical developments.

2. RESEARCH METHODS

2.1 Data Source and Search Strategy

The data for this study were obtained from Scopus, which is widely recognized as one of the largest and most authoritative sources of peer-reviewed literature across multiple disciplines. Scopus was selected due to its extensive coverage, high-quality indexing, and strong compatibility with bibliometric analysis tools. A structured search query was employed to capture publications related to federated learning and

distributed computing, incorporating keyword combinations such as “federated learning,” “distributed computing,” “edge computing,” and “privacy-preserving machine learning.” The search was restricted to titles, abstracts, and keywords to ensure the relevance and precision of the dataset. Furthermore, the analysis spans publications from 2010 to 2025, reflecting both the early emergence and the rapid expansion of federated learning research within distributed computational environments.

2.2 Data Collection and Screening

The initial search results were exported from Scopus in CSV format, including key bibliographic information such as author names, publication titles, abstracts, keywords, affiliations, citation counts, and publication years. To ensure data quality and reliability, a rigorous screening process was conducted to eliminate duplicate records, incomplete entries, and irrelevant publications that did not directly address federated learning or its application within distributed computing contexts. The inclusion criteria comprised peer-reviewed journal articles and conference proceedings, publications written in English, and studies explicitly focusing on federated learning, distributed systems, or closely related topics. Conversely, exclusion criteria encompassed non-scholarly documents such as editorials and notes, studies from unrelated fields, and publications lacking sufficient bibliographic information for meaningful analysis.

2.3 Bibliometric Analysis Techniques

This study applies two principal bibliometric techniques—performance analysis and science mapping—to comprehensively examine the research landscape. Performance analysis is employed to evaluate research productivity and scholarly impact, encompassing metrics such as annual

publication trends, the most prolific authors, leading institutions, and contributing countries, alongside citation analysis to identify the most influential publications within the field. Complementarily, science mapping is utilized to explore the intellectual structure and relational dynamics of the research domain, including co-authorship analysis to assess collaboration networks among researchers, institutions, and countries; keyword co-occurrence analysis to identify dominant research themes and emerging topics; and citation as well as co-citation analysis to uncover the foundational literature and knowledge clusters that shape the development of the field.

2.4 Data Visualization and Analysis Tools

To facilitate network visualization and clustering, this study employs VOSviewer, which enables the construction of bibliometric maps based on co-authorship, keyword co-occurrence, and citation relationships [10], [11]. Through its clustering algorithm, VOSviewer groups related items into distinct thematic clusters that are visualized using network, overlay, and density maps. In addition, descriptive statistical analysis is conducted to illustrate trends in publication output, citation growth, and distribution across subject areas, thereby providing a quantitative basis for understanding the evolution of federated learning research. Overall, the research framework is structured into four main stages: data retrieval from Scopus, data cleaning and screening, bibliometric analysis using performance and science mapping techniques, and interpretation of results. This systematic approach ensures the reliability and validity of the findings while offering a comprehensive overview of the research landscape.

3. RESULTS AND DISCUSSION

3.1 Author Collaboration Analysis

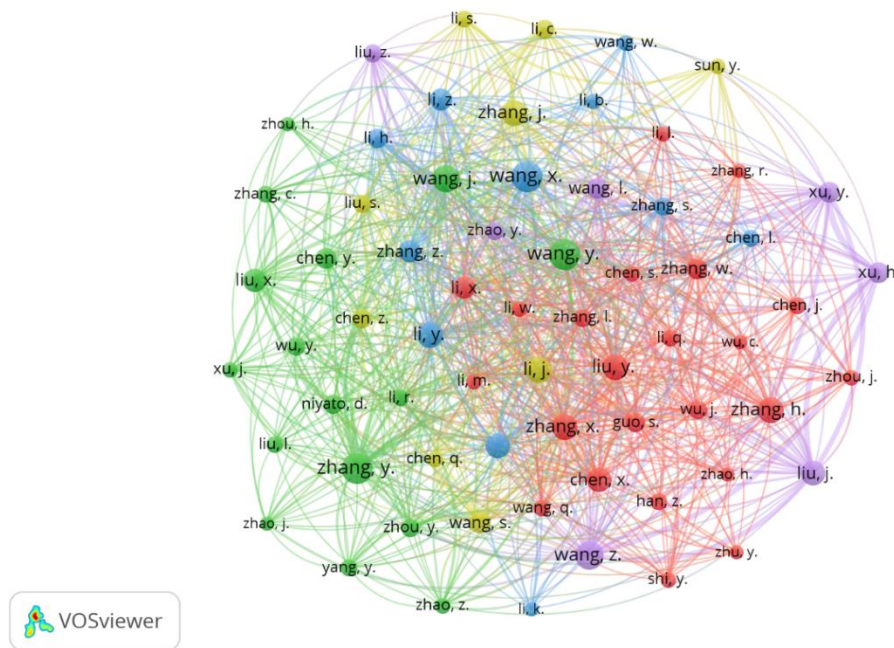


Figure 1. Author Visualization
Source: Data Analysis

Figure 1 illustrates the co-authorship network of researchers in federated learning using VOSviewer, showing a densely connected structure with multiple-colored clusters that represent distinct research collaboration groups. The visualization indicates that the field is highly collaborative, with several dominant authors—such as Wang, Zhang, Liu, and Chen—occupying central positions in the network, suggesting their strong influence and productivity. The close proximity and numerous links among

nodes reflect frequent co-authorship and active knowledge exchange, while the presence of inter-cluster connections highlights cross-group and international collaboration. Overall, this pattern demonstrates that federated learning research in distributed computing is supported by a robust and globally interconnected scholarly community, which plays a crucial role in accelerating innovation and advancing the field.

3.2 Citation Analysis: Influential Publications

Table 1. Most Cited Article

Citations	Author and Year	Title	Publication
1949	[12]	Adaptive Federated Learning in Resource Constrained Edge Computing Systems	IEEE Journal on Selected Areas in Communications
1559	[13]	Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge	IEEE International Conference on Communications
1104	[14]	Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT	IEEE Transactions on Industrial Informatics

Citations	Author and Year	Title	Publication
1006	[15]	Federated Learning over Wireless Networks: Optimization Model Design and Analysis	Proceedings - IEEE INFOCOM
981	[16]	Federated learning via over-the-air computation	IEEE Transactions on Wireless Communications
765	[17]	Swarm Learning for decentralized and confidential clinical machine learning	Nature
694	[18]	Broadband Analog Aggregation for Low-Latency Federated Edge Learning	IEEE Transactions on Wireless Communications
683	[19]	A Survey on Federated Learning for Resource-Constrained IoT Devices	IEEE Internet of Things Journal
571	[20]	Edge Learning for B5G Networks With Distributed Signal Processing: Semantic Communication, Edge Computing, and Wireless Sensing	IEEE Journal on Selected Topics in Signal Processing
469	[21]	Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks	IEEE Transactions on Industrial Informatics

Source: Scopus, 2026

Table 1 shows that the most highly cited articles in this field are dominated by studies that connect federated learning with edge computing, wireless communication, IoT, blockchain, and privacy-preserving systems, indicating that the research agenda has been strongly shaped by practical implementation issues rather than purely theoretical discussion. The most cited article, Adaptive Federated Learning in Resource Constrained Edge Computing Systems by Wang et al., with 1,949 citations, highlights the major concern of adapting federated learning to limited computational environments, while other highly cited works such as Nishio and Yonetani on client selection, Tran et al. on wireless networks, and Yang et al. on over-the-air computation confirm that efficiency, resource heterogeneity, and communication optimization are central

themes in the literature. In addition, the strong citation impact of Lu et al.'s studies on blockchain integration and Warnat-Herresthal et al.'s work on decentralized clinical machine learning in Nature suggests that security, trust, and real-world application in sensitive sectors such as healthcare and industrial IoT have significantly expanded the relevance of federated learning research. Overall, this citation pattern demonstrates that the intellectual foundation of federated learning in distributed computing is built around solving distributed data privacy problems while simultaneously addressing latency, scalability, and deployment challenges in modern networked systems.

3.3 Keyword Co-Occurrence and Research Themes

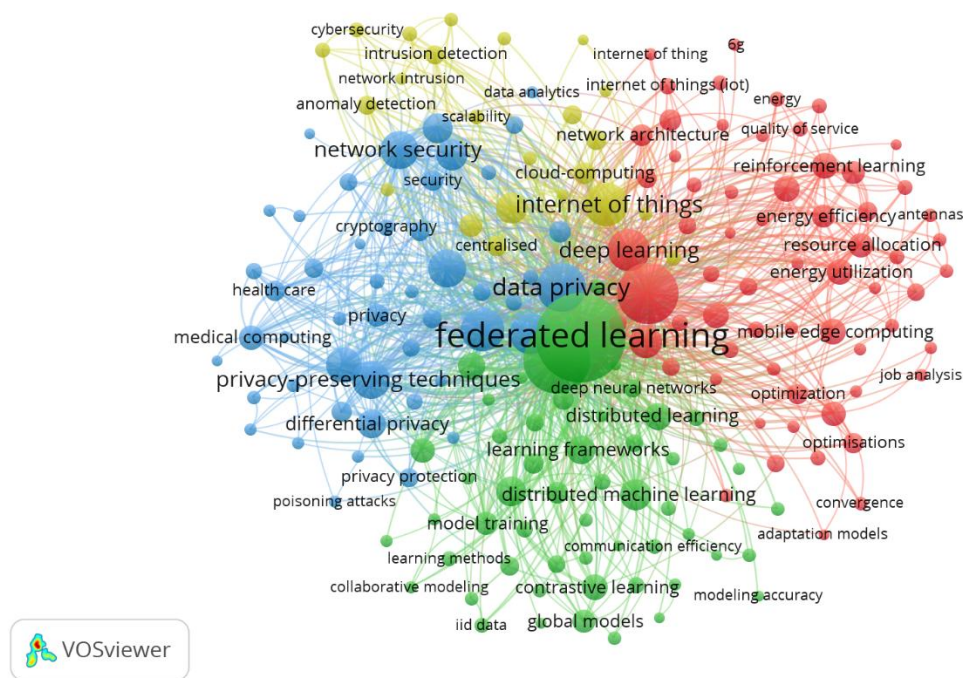


Figure 2. Network Visualization
Source: Data Analysis

Figure 2 presents the keyword co-occurrence network generated using VOSviewer, which highlights the main research themes and their interrelationships in federated learning studies. The visualization reveals several prominent clusters, with “federated learning” positioned as the central node, indicating its role as the core concept connecting various research streams. The red cluster is primarily associated with optimization, resource allocation, mobile edge computing, and energy efficiency, reflecting technical challenges in system performance and scalability. The blue cluster emphasizes privacy, security, and cryptographic techniques such as differential privacy and privacy-preserving methods, showing the strong

focus on data protection. Meanwhile, the green cluster relates to distributed machine learning, model training, and communication efficiency, indicating foundational aspects of collaborative learning systems. The yellow cluster, centered around Internet of Things (IoT), network architecture, and cloud computing, highlights the application context of federated learning. Overall, this network demonstrates that federated learning research is highly interdisciplinary, integrating concerns of privacy, system optimization, and real-world deployment, with increasing convergence toward scalable and secure implementations in distributed environments.

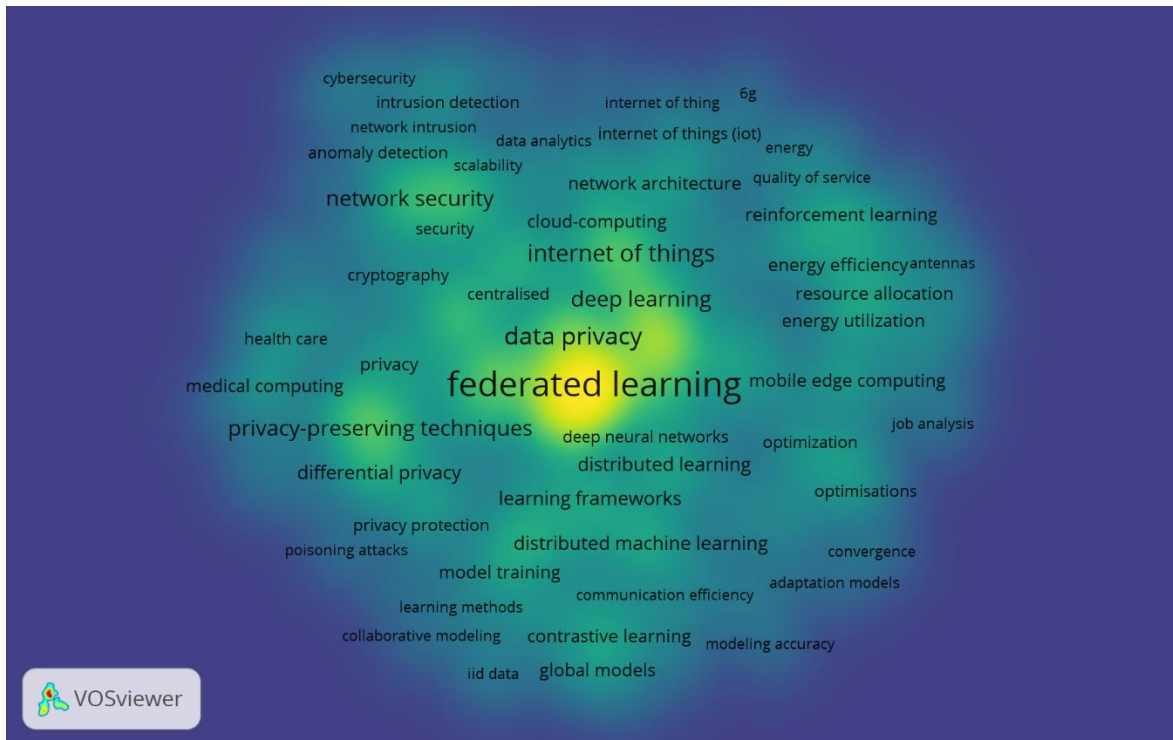


Figure 4. Density Visualization
Source: Data Analysis

Figure 4 presents the density visualization generated using VOSviewer, which illustrates the concentration and intensity of research topics within the field of federated learning. The color gradient—from blue (low density) to yellow (high density)—indicates the frequency and prominence of keywords. The most intense (yellow) area is centered around “federated learning,” confirming its position as the core focus of the research domain. Surrounding high-density areas include “data privacy,” “deep learning,” and “internet of things,” highlighting these as the most frequently studied and strongly interconnected themes. Medium-density zones (green) represent supporting topics such as distributed machine learning, model training, and communication efficiency, while lower-density areas (blue) include more specialized or emerging topics like contrastive learning, global models, and poisoning attacks. Overall, the density map indicates that research in federated learning is heavily concentrated on

privacy-preserving mechanisms and real-world applications, while also expanding into optimization and advanced machine learning techniques, reflecting both the maturity and ongoing diversification of the field.

3.4 Discussion

The findings of this bibliometric analysis demonstrate that federated learning has evolved into a central paradigm within distributed computing, driven primarily by the increasing need for privacy-preserving data processing and decentralized intelligence. The rapid growth in publications, particularly after 2018, reflects a strong global response to challenges associated with data security, regulatory compliance, and the limitations of centralized machine learning models. The dominance of highly cited works focusing on edge computing, wireless networks, and IoT environments indicates that the field is not only theoretically driven but also highly application-oriented, addressing real-world constraints such as limited

resources, latency, and communication efficiency [22]–[24].

The co-authorship and collaboration network analysis further reveals that federated learning research is supported by a globally interconnected academic and industrial community. Strong collaboration patterns across countries such as China, the United States, and Europe highlight the strategic importance of this field in national and international research agendas. The presence of key authors acting as central nodes suggests that knowledge diffusion is facilitated by influential researchers who bridge multiple collaboration clusters. This interconnected structure accelerates innovation and promotes interdisciplinary integration, combining expertise from machine learning, network engineering, cybersecurity, and domain-specific applications [25]–[27].

From a thematic perspective, the keyword co-occurrence and density analyses indicate that research in federated learning is heavily concentrated around core issues such as data privacy, deep learning, and IoT integration. These themes form the backbone of the field, while emerging topics such as energy efficiency, reinforcement learning, blockchain integration, and mobile edge computing reflect a shift toward optimization and scalability. The overlay visualization further confirms this evolution, showing a transition from foundational studies on distributed learning and model training to more advanced and applied research focusing on system performance and real-world deployment.

Despite these advancements, several challenges remain that limit the widespread adoption of federated learning. Issues such as data heterogeneity, model convergence instability, communication overhead, and vulnerability to adversarial attacks continue to pose significant technical barriers. Additionally, while federated learning addresses many privacy

concerns, it does not fully eliminate risks related to inference attacks or data leakage through model updates. These challenges suggest that future research should focus on developing more robust, secure, and efficient algorithms, as well as standardized frameworks for implementation.

In addition, the integration of federated learning with emerging technologies presents both opportunities and complexities. The combination with blockchain offers enhanced transparency and trust, while the use of edge computing and 5G/6G networks enables real-time and large-scale deployment. However, these integrations also introduce new layers of system complexity and require careful consideration of trade-offs between performance, security, and resource utilization. Therefore, interdisciplinary approaches and collaborative research efforts will be essential in addressing these multifaceted challenges.

Overall, the discussion highlights that federated learning is transitioning from an emerging concept to a mature and widely applicable technology within distributed computing. The field is characterized by rapid growth, strong collaboration, and increasing diversification of research themes. Moving forward, the focus is expected to shift toward practical implementation, standardization, and addressing unresolved challenges, ensuring that federated learning can be effectively deployed in real-world, large-scale, and privacy-sensitive environments.

4. CONCLUSION

This study provides a comprehensive bibliometric overview of federated learning research within the context of distributed computing, highlighting its rapid growth, evolving research focus, and increasing global relevance. The findings indicate that federated learning has transitioned from a conceptual framework into a widely

explored and applied research domain, driven by the need for privacy-preserving and decentralized data processing. The analysis reveals that key research themes are centered on data privacy, deep learning, and Internet of Things (IoT) applications, while emerging topics such as energy efficiency, blockchain integration, and resource optimization reflect the field's progression toward practical implementation. Furthermore, collaboration analysis demonstrates that federated learning research is supported by a strong and interconnected global network of researchers and institutions, facilitating knowledge exchange and accelerating innovation.

In addition, citation analysis confirms that highly influential studies primarily focus on addressing critical technical challenges, including communication efficiency, resource

constraints, and system scalability, underscoring the importance of interdisciplinary approaches in advancing the field. Despite these advancements, several challenges remain, such as data heterogeneity, model convergence issues, communication overhead, and security vulnerabilities, which must be addressed to enable large-scale and real-world deployment. Future research should therefore prioritize the development of more robust algorithms, enhanced system efficiency, and standardized frameworks to support broader adoption. Overall, federated learning is positioned as a key enabler of next-generation distributed computing systems, playing a crucial role in balancing advanced data analytics with the growing demands for privacy, security, and regulatory compliance in an increasingly data-driven world.

REFERENCES

- [1] F. Yin, Z. Lin, Q. Kong, Y. Xu, D. Li, and ..., "FedLoc: Federated learning framework for data-driven cooperative localization and location data processing," *IEEE Open J. ...*, 2020.
- [2] X. Wan, D. Yang, T. Wang, and M. Deveci, "Closed-loop supply chain decision considering information reliability and security: should the supply chain adopt federated learning decision support systems?," *Ann. Oper. Res.*, pp. 1–37, 2023.
- [3] M. Chekour, N. El Morabit, I. Benqassou, A. Lechhab, F. El-Hars, and M. M. Hafid, "Integrating Blended Learning in Teacher Training: A Comprehensive Review," in *2024 Mediterranean Smart Cities Conference (MSCC)*, 2024, pp. 1–4.
- [4] R. Dogra, H. Babbar, and S. Rani, "Integration of WSN and IoT: Its Applications and Technologies," in *IoT and WSN based Smart Cities: A Machine Learning Perspective*, Springer, 2022, pp. 243–256.
- [5] W. Dhewanto, A. N. Umbara, and R. Hanifan, "Towards Policy Development of Entrepreneurial Ecosystem: A Review in Indonesia Financial Technology Sector," in *Proceedings of the 8th International Conference on Industrial and Business Engineering*, 2023, pp. 282–290. doi: 10.1145/3568834.3568841.
- [6] S. Kovalyov, "An approach to develop a generative design technology for power systems," in *Vlth International Workshop'Critical Infrastructures: Contingency Management, Intelligent, Agent-Based, Cloud Computing and Cyber Security'(IWCI 2019)*, 2019, pp. 79–82.
- [7] R.-M. Dumitrescu, "Processing of personal and medical data by judicial institutions in the context of the enforcement of Regulation EU 2016/679–General Data Protection Regulation (GDPR)," *J. Comp. Res. Anthropol. Sociol.*, vol. 9, no. 01, pp. 1–18, 2018.
- [8] E. S. Dove, "The EU general data protection regulation: implications for international scientific research in the digital era," *J. Law, Med. Ethics*, vol. 46, no. 4, pp. 1013–1030, 2018.
- [9] M. Goddard, "The EU General Data Protection Regulation (GDPR): European regulation that has a global impact," *Int. J. Mark. Res.*, vol. 59, no. 6, pp. 703–705, 2017.
- [10] F. C. Fenerich, K. Guedes, N. H. M. Cordeiro, G. de S. Lima, and A. L. G. De Oliveira, "Energy efficiency in industrial environments: an updated review and a new research agenda," *Rev. Gestão e Secr. (Management Adm. Prof. Rev.)*, vol. 14, no. 3, pp. 3319–3347, 2023, doi: 10.7769/gesec.v14i3.1802.
- [11] B. Harsanto, "Innovation management in the library: A bibliometric analysis," *Libr. Philos. Pract.*, vol. 5908, 2021.
- [12] S. Wang et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. areas Commun.*, vol. 37, no. 6, pp. 1205–1221, 2019.
- [13] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE international conference on communications (ICC)*, 2019, pp. 1–7.
- [14] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Trans. Ind. Informatics*, vol. 16, no. 6, pp. 4177–4186, 2019.
- [15] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *IEEE INFOCOM 2019-IEEE conference on computer communications*,

- 2019, pp. 1387–1395.
- [16] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 3, pp. 2022–2035, 2020.
- [17] S. Warnat-Herresthal *et al.*, "Swarm learning for decentralized and confidential clinical machine learning," *Nature*, vol. 594, no. 7862, pp. 265–270, 2021.
- [18] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 1, pp. 491–506, 2019.
- [19] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained IoT devices," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 1–24, 2021.
- [20] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Top. Signal Process.*, vol. 17, no. 1, pp. 9–39, 2023.
- [21] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks," *IEEE Trans. Ind. Informatics*, vol. 17, no. 7, pp. 5098–5107, 2020.
- [22] I. P. Chochliouros, M. A. Kourtis, A. S. Spiliopoulou, and ..., "Energy efficiency concerns and trends in future 5G network infrastructures," *Energies*, 2021.
- [23] A. Chawla *et al.*, "IoT-Based Monitoring in Carbon Capture and Storage Systems," *IEEE Internet Things Mag.*, vol. 5, no. 4, pp. 106–111, 2022.
- [24] H. Jain, V. Chamola, and Y. Jain, "5G network slice for digital real-time healthcare system powered by network data analytics," *Internet of Things and Cyber-Physical ... Elsevier*, 2021.
- [25] S. Sood and A. Kim, "The Golden Age of the Big Data Audit: Agile Practices and Innovations for E-Commerce, Post-Quantum Cryptography, Psychosocial Hazards, Artificial Intelligence Algorithm Audits, and Deepfakes," *Int. J. Innov. Econ. Dev.*, vol. 9, no. 2, pp. 7–23, 2023, doi: 10.18775/ijied.1849-7551-7020.2015.92.2001.
- [26] P. J. Nesse, H. S. Hallingby, and ..., "Validation of 5G use case solutions—Simultaneous assessment of business value and social acceptance in early stages of the research and innovation ...," *... and Technology. journal.riverpublishers.com*, 2023.
- [27] A. T. Oyewole, C. C. Okoye, O. C. Ofodile, and C. E. Ugochukwu, "Cybersecurity risks in online banking: A detailed review and preventive strategies applicatio," *World J. Adv. Res. Rev.*, vol. 21, no. 3, pp. 625–643, 2024.