**West Science Information System and Technology**
Vol. 4, No. 01, April 2026, pp. 112~122

# Data Privacy Studies (2010–2026): A Scopus-Based Bibliometric Analysis of Research Hotspots and Citation Dynamics

**Loso Judijanto**
IPOSS Jakarta, Indonesia

**Article Info**

***Article history:***

Received Apr, 2026
Revised Apr, 2026
Accepted Apr, 2026

***Keywords:***

Data Privacy
Bibliometric Analysis
Scopus
VOSviewer
Research Hotspots

**ABSTRACT**

The current study is designed to analyze the research on data privacy between the years 2010 and 2026 using bibliometrics. It seeks to explore collaboration trends, influential literature, and emerging research trends within the research area through co-authorship, citations, and co-keywords analysis. VOSViewer was used for the analysis, as well as visualization of co-citation and co-keyword networks, thus mapping out the intellectual and conceptual network of data privacy. As can be seen from the results, there have been many developments regarding the field under review, marked by increased international collaboration, especially among top-ranking countries such as China, the United Kingdom, India, and Germany. The citation network shows that the field relies not only on technologies like differential privacy and machine learning but also on behavior-based factors such as trust and privacy. The keyword network further indicates that "data privacy" plays an important role in this research domain, together with artificial intelligence, deep learning, and federated learning. The change in time for keywords implies a transition from traditional security methods to privacy-oriented techniques.



***Corresponding Author:***

Name: Loso Judijanto
Institution: IPOSS Jakarta, Indonesia
Email: losojudijantobumn@gmail.com

## 1. INTRODUCTION

The rapid development and increase in data creation in the modern digital age have elevated data to an important resource within different industries. Due to the development of different digital technologies such as cloud computing, AI, and mobile apps, there has been an increasing emphasis on the accumulation, storage, and analysis of personal data. This phenomenon has made the issue of data privacy a growing concern not only within academic circles but also from the policymaking perspective [1].

In the last decade, there has been a significant increase in the study of data privacy, which has been influenced by the rise in data breaches, changes in legal systems, and heightened public awareness. There has been a substantial increase in the publication of studies concerning the field of privacy, especially in recent years since 2015 when the digitalization of businesses reached a high peak. The trend in increasing academic publications highlights the need for studying data privacy within multiple disciplines, including computer science, law, business, and social sciences [2], [3].

The bibliometric method has been recognized as a promising tool that can be used to study scientific knowledge and measure research efficiency. Using a set of publication metadata like author affiliation, references, keywords, and co-authorship relations, bibliometrics studies can offer useful information on how a particular field of research evolved. Scopus database has proven useful in carrying out such studies thanks to its large body of scientific literature sources in various formats, like journal papers, proceedings articles, and book sections . Such an inclusive database provides a much broader perspective on the global picture of research, especially on contributions made by new countries and cross-disciplinary areas [4], [5].

It is evident from previous bibliometrics carried out within different fields that an increase in publication rate as well as the evolution of the thematic content and collaboration networks over time exists in each case. For example, certain fields like artificial intelligence, sustainability, and financial inclusion have shown great growth in the last ten years and have also changed their theme and methodology over time. Likewise, studies that deal with issues related to privacy have also depicted the dynamic nature of this field. New themes in this context include mobile privacy, legislation related to data protection, and user trust issues among others.

However, although a large amount of literature is currently available, it is necessary to conduct an up-to-date synthesis of the scientific literature on the topic, which would take into account not only current research areas but also their citations over time. Current papers tend to be limited either by their subject area or by the period under consideration, thereby depriving researchers of a general view of how the field evolved. Another important aspect of any bibliometric study is the citation analysis, which can help researchers understand the influence of previous works on the further development of privacy science.

While data privacy has evolved into an important topic of investigation with a growing body of scholarly works, there is a noticeable absence of a bibliometric study of the topic that provides an integrated view by incorporating its publication history, topic changes, and citation behaviors over time. The previous analyses on data privacy tend to be narrowly focused on particular fields, such as the application of mobile devices and consumer privacy issues, or they cover only a few years. Moreover, inconsistencies among databases and methodologies can cause fragmented results, preventing researchers from forming a holistic perspective on global research trends. Consequently, it is essential to conduct an extensive analysis using a well-established and robust dataset to trace the evolution of data privacy studies between 2010 and 2026.

The present study attempts to perform a bibliometric analysis of data privacy studies listed in the Scopus database from 2010 to 2026. The specific objectives of this research include (1) analyzing publication trends during the selected time period; (2) identifying research themes and hotspots using keyword and co-occurrence analysis; and (3) conducting a citation analysis to identify the authors, organizations, and sources that have had a significant influence on the field of data privacy. A thorough examination of the intellectual structure of data privacy studies is expected to shed light on important developments in this field.

## 2. METHODS

A bibliometric approach that falls under the scope of quantitative analysis is used to examine systematically the development of research in data privacy from 2010 to 2026. A bibliometric analysis is known to be a highly efficient means of assessing scientific productivity and trends, and mapping out the conceptual framework of a certain discipline. Data used for this study were collected exclusively via searches performed in the Scopus database due to its comprehensive collection of high-caliber peer-reviewed journals in various disciplines. An advanced search string was designed based on appropriate keywords including

"data privacy," "information privacy," and "personal data protection," which was applied to title/abstract/key words fields. This particular search strategy was further restricted by setting time parameters between 2010 and 2026 and specifying document types.

The bibliometric data that were extracted from the database were subsequently processed and analyzed with the use of bibliometric software such as VOSviewer and Excel. Besides thematic analysis, citation analysis was undertaken to determine the impact and influences of publications in the corpus of works analyzed. Factors such as number of citations received by each publication, number of citations per document, as well as number of highly cited papers were considered to find influential papers and influential authors. Moreover, co-citation analysis and bibliographic coupling analysis were carried out to analyze the intellectual structure and knowledge domain of data privacy.

## 3. RESULTS AND DISCUSSION

### 3.1 Co-Authorship Analysis

The technique of co-authorship analysis is used to investigate the nature of collaboration in data privacy research through the identification of links between authors, organizations, and countries. The findings of co-authorship analysis offer insight into patterns of scientific collaboration, emphasizing important contributors and knowledge transfer in the discipline. Through co-authorship network mapping, the study explores dominant areas of research and their interconnectivity in the development of data privacy research between 2010 and 2026.

1. Author-level Visualization

The co-authorship network diagram produced via VOSviewer highlights the pattern of collaborations between authors working on data privacy research. This diagram reveals how authors collaborate based on common publications, which creates clusters representing collaborations or communities of researchers. The positioning of authors in the diagram, their coloring, and the connecting lines offer information on the level of collaboration, clustering of research, and the importance of certain authors in connecting other authors within the collaboration network.

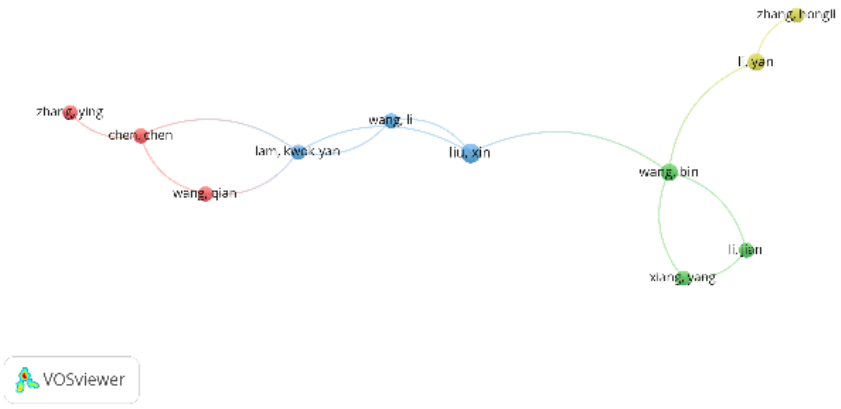


Figure 1. Author-level Visualization

*Source: Data Analysis*

The figure demonstrates that there exist a number of clusters in the graph that contain sets of authors who interact with each other quite often. Different colors show clusters, which signify research groups in the area of data privacy, which operate independently from each other to some extent. Thus, the red cluster on the left hand

side contains authors who interact closely among themselves but have no connections with authors in other clusters. The same can be said about the green cluster on the right hand side of the graph.

The case is quite opposite when it comes to the blue cluster in the middle of the map. The authors in this cluster seem to be in a connecting position since they occupy a position among other clusters and are also connected to each other through various links. Thus, the authors in this blue cluster can act as knowledge brokers in the research environment.

2. Institution-level Visualization

Institutional-level authorship network visualizes the collaboration among different universities and research institutions involved in conducting research on data privacy. Using the VOSViewer software, the figure illustrates the interconnectedness of these institutions based on research collaboration output. By looking at clusters, nodes, and their strength, one can understand more about the institutions that collaborate with each other the most, as well as the volume of knowledge exchange between different academic institutions.

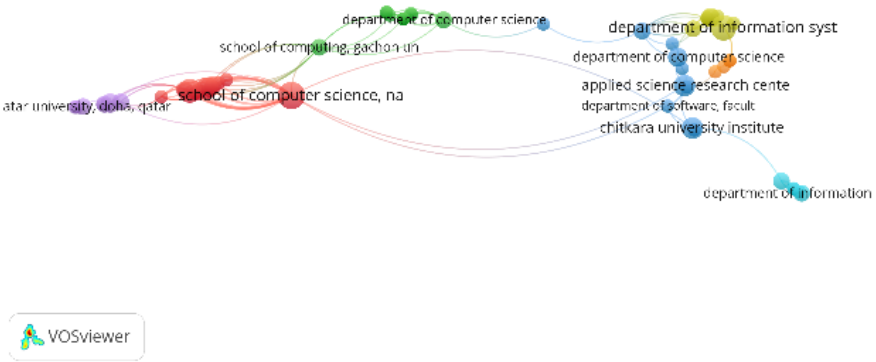


Figure 2. Institution-level Visualization

*Source: Data Analysis*

Institutional clusters can be clearly seen in the diagram, with each of the clusters comprising university partners that regularly work together within the network. For example, the clusters marked red and purple on the left-hand side of the diagram represent institutions that have strong linkages, such as those between Qatar University and its constituent departments, thus indicating high levels of collaboration within a region or within an institution. The same applies to the green cluster, which comprises institutions including Soochow University, and represents a collaboration network within a localized area.

The right-hand side of the chart shows a different variety of institutions with computer science and information systems departments. The yellow and light-blue clusters point to a greater degree of collaborative activity with an interdisciplinary approach. For instance, institutions like Chitkara University Institute and different information systems departments show multiple links with each other, which reflects institutional interaction with one another. The connecting links between these clusters indicate emerging cooperation on both regional and international levels, even if data privacy research within institutions remains localized to some extent.

3. Country-level Visualization

The country co-authorship network analysis graph provides insights on the trends

in terms of international collaborations for data privacy studies between 2010 and 2026. This visual graph created using VOSviewer shows the connections between different countries in relation to the publications they have produced together, emphasizing their collaboration level in their scientific research endeavors. Nodes in this graph show the significance of different countries as contributors, whereas the links show the connections between different countries.

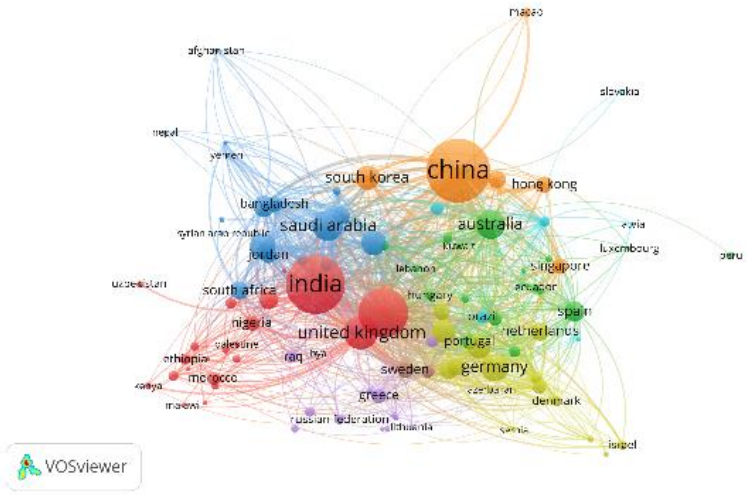


Figure 3. Country-level Visualization
*Source: Data Analysis*

From the visualization, it is clear that data privacy studies have very high internationalization levels, with many leading countries serving as major nodes in the network of collaboration. China, for instance, emerges as one of the major nodes, thereby showing its contribution to the study and its active involvement in international collaborations with other nations. Likewise, nations like India, the UK, and Germany emerge as major nodes in their clusters, implying that they are productive in terms of research and collaborate actively with other countries internationally.

As can be seen from the figure above, there is very high internationalization among research on data privacy, with most of the leading countries playing the role of nodes in the network of collaboration. For example, China comes out as one of the major nodes, hence indicating the contribution of the country towards the field of research and its level of collaboration with other nations. Similarly, countries such as India, the United Kingdom, and Germany come out as major nodes in their respective clusters.

### 3.2 Citation Analysis

Citation analysis is conducted to evaluate the intellectual influence and impact of publications within the data privacy research domain.

Table 1. The Most Impactful Literatures

| Citations | Authors and year | Title |
|---|---|---|
| 7189 | [6] | Edge Computing: Vision and Challenges |
| 6197 | [7] | Calibrating noise to sensitivity in private data analysis |
| 4444 | [8] | Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon |
| 2901 | [9] | The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes |

| Citations | Authors and year | Title |
|---|---|---|
| 2737 | [10] | A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents |
| 2697 | [11] | Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model |
| 2576 | [12] | MixMatch: A holistic approach to semi-supervised learning |
| 2562 | [13] | Model inversion attacks that exploit confidence information and basic countermeasures |
| 2466 | [14] | Big Data in Smart Farming – A review |
| 2293 | [15] | Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective |

*Source: Scopus, 2026*

The papers mentioned in Table 1 are the most impactful literature sources forming the theoretical background of data privacy, as well as other relevant areas, due to their high citations. This suggests that technological innovations and privacy-preserving methods are two cornerstones of the area under analysis. At the same time, it is noteworthy that several seminal studies focus on user behavior and trust perceptions, including the works by Malhotra et al. (2004) and Kim et al. (2008). As such, one can suggest that behavioral and organizational theories play an integral role in data privacy analysis. Lastly, it is important to note that there are a number of papers devoted to such topical issues as big data (e.g., Boyd & Crawford, 2012; Wolfert et al., 2017), as well as new security threats, such as model inversion attacks (e.g., Fredrikson et al., 2015).

### 3.3 Keyword Co-Occurrence Analysis

Co-occurrence analysis of keywords is applied to uncover the theoretical framework and theme progression in the studies on data privacy. This method makes it possible to determine the main topics in the field, as well as emerging themes and changes in research focus over time, based on the frequency of keywords and their co-occurrence.

1. Network Visualization

Network visualization for keyword co-occurrences is a visual representation that gives an overview of the conceptual structure and development of data privacy studies. This network diagram has been created using the software application called VOSviewer and shows the connections between commonly used keywords. Each point in this map is a particular keyword, and links denote co-occurrences. Different colors show the different main topics studied in the field of data privacy studies, and the closeness of points to each other denotes the degree of relationship between various concepts.

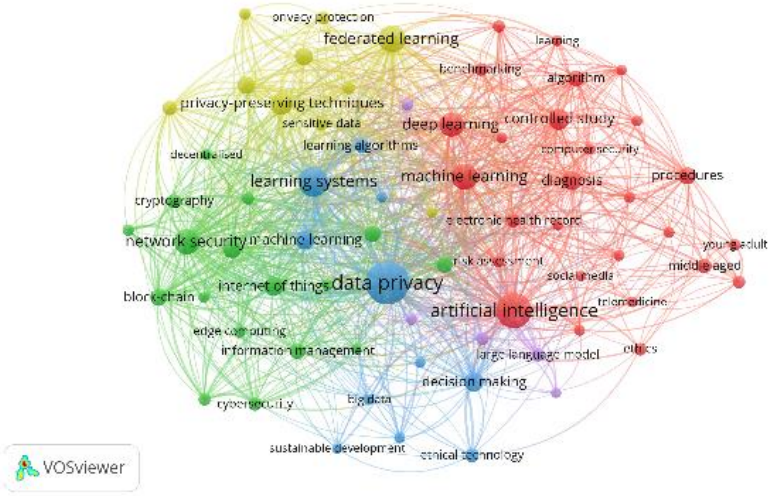


Figure 4. Network Visualization
*Source: Data Analysis*

The visual shows a few highly connected clusters, which altogether form the cognitive map of data privacy research. The primary topic of focus is "data privacy," which forms a core concept that is linked to many technical and technological terms. "Machine learning," "deep learning," and "artificial intelligence" are some of the closely connected keywords, which implies that the study of data privacy is becoming an essential part of computational technologies. This implies that contemporary issues of privacy are significantly influenced by the proliferation of intelligent machines, which require sophisticated data processing methods.

One of the other key clusters focuses on privacy-enhancing techniques. This includes keywords like “federated learning,” “privacy protection,” and “privacy-enhancing techniques.” It is interesting to note that there is an increasing trend towards developing technologies that would help utilize data without compromising privacy. The inclusion of “federated learning” in the list of keywords indicates that research is increasingly adopting a decentralized model of processing data. Here, data remains in its original form and is processed collectively using the models.

Thirdly, there is a cluster which centers on security-related themes, where the search terms include “cryptography,” “network security,” “blockchain,” and “cybersecurity.” The presence of these terms suggests that data privacy is linked closely to the issue of information security in general. The involvement of blockchain technology, in particular, suggests that there will be new ways devised to ensure data integrity and transparency, whereas cryptographic techniques play an important role in securing sensitive information.

Furthermore, the clustering shows that another group involves issues pertaining to decision-making and governance, such as the terms “decision making,” “controlled study,” and “procedures.” In this way, the research on data privacy does not only focus on solving problems through technological means but also tackles organizational issues. The fact that there are keywords involving assessment or evaluation reveals that there is an increased interest in evaluating the success of various privacy measures.

Keywords related to “emerging” and “interdisciplinary” fields like “ethical technology,” “sustainable development,” and “big data” show that the subject area of data privacy is becoming increasingly broader. It seems that the research is moving towards a more holistic outlook, one which not only addresses issues relating to data but integrates considerations of ethics and sustainability into the discussion.

2. Overlay Visualization

The co-occurrence keyword overlay visualization depicts the development of research themes in relation to the history of the research into data privacy. With the use of VOSviewer, this visualization tool not only depicts the connections between the different keywords but also shows their evolution with respect to time in terms of colors used. The darker the keyword, the older the research focus; the brighter the color, the newer the topic being investigated.

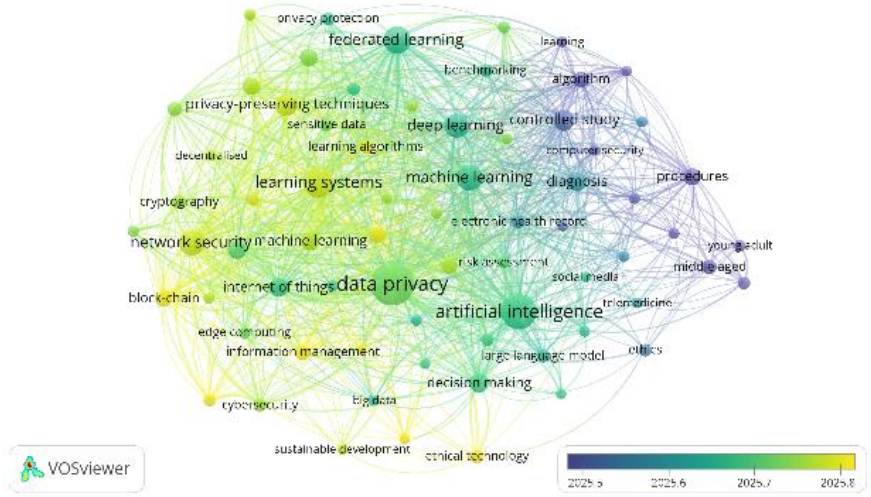


Figure 5. Overlay Visualization

*Source: Data Analysis*

As can be seen in the graph above, prior research conducted in the field of data privacy was mainly related to basic principles and theories, including "data privacy," "learning systems," and "machine learning," as shown through their comparatively dark color shade. This indicates that prior work in the area concentrated on laying down the basic connection between the processing system and data privacy. Besides, the inclusion of phrases such as "cryptography" and "network security" in early stages highlights that prior work was based on conventional security principles.

In the course of time, there have been some changes as far as computational aspects are concerned. There has been an increase in the importance of using highly computational algorithms like "deep learning" and "artificial intelligence." Keywords that show this trend can be seen in mid-color tones since they show the increasing significance over time. This change shows how advancements in technology in the form of AI have changed the nature of privacy concerns and require new methods to be employed in order to handle them.

Trends from more recent years, marked with brighter colors, show the development of more specialized and future-oriented concepts like "federated learning," "privacy-preserving methods," and "ethical technologies." These search terms suggest that the discussion has moved toward developing technologies that improve privacy without reducing the usability of the data. The increasing emphasis on ethics and sustainability shows that the debate is not only technological but social as well.

3. Density Visualization

The density graph based on keyword co-occurrence is a concentrated view of the intensity and density of research topics in the field of data privacy research. Through VOSviewer software, this visual presentation points out areas having a high occurrence of keywords; bright areas mean that the area represents a high number of research activities while dark areas indicate under-researched areas.

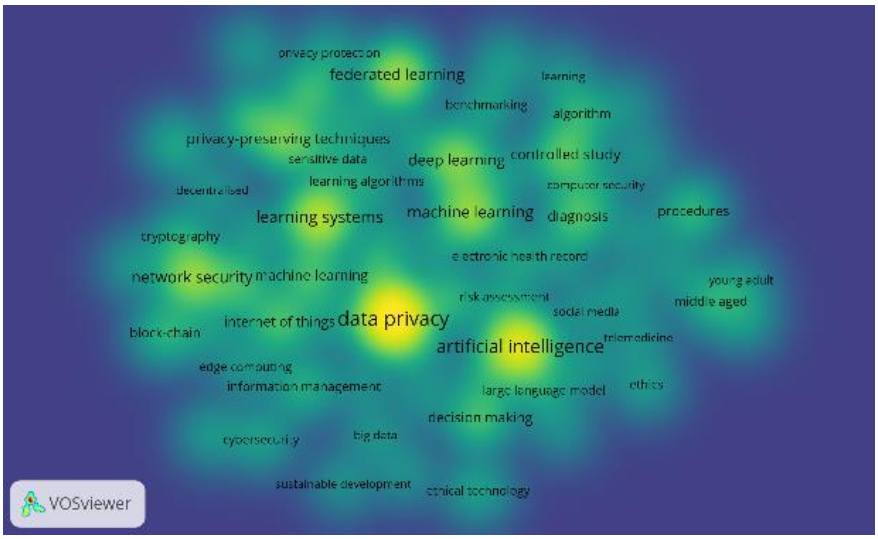


Figure 6. Density Visualization
Source: Data Analysis

It is quite clear from the diagram above that the word "data privacy" acts as the central theme of all the keywords, which further signifies that it is the dominating topic in the field of research due to its concentration at the center. Moreover, some other high-density keywords such as "artificial intelligence," "machine learning," and "deep learning," along with the central topic, reveal the fact that a considerable number of studies in recent years have been conducted on combining privacy with advanced computing.

Outside of the core cluster, there are other moderately concentrated clusters that contain key terms including “privacy-preserving methods,” “federated learning,” “learning frameworks,” and “network security.” These denote active research domains that further contribute to the main field of data privacy. On the other hand, there are lesser concentrated clusters that consist of terms such as “ethics in technology,” “sustainable development,” and “cybersecurity.”

**Discussion**

From the results presented above, we can conclude that the study of data privacy is now a very lively field that is constantly developing due to technological innovations and rising public interest in protecting personal information. The analysis of co-authorship shows that there are separate and isolated groups of authors, organizations, and countries, but they have been increasingly interacting with each other. However, despite the widespread activity of research production, the field is dependent on certain centers, including such countries as China, the United Kingdom, India, and Germany, in fostering international cooperation. In this way, we observe the general tendency of scientific research, when the most developed nations function as knowledge brokers.

From an intellectual point of view, the citation analysis shows that the basis for research in data privacy is intrinsically linked to both technological and behavioral paradigms. Top papers about differential privacy, edge computing, and machine learning security show that technology-driven approaches are key in solving privacy problems. On the other hand, top research on trust, perceived risk, and user privacy concerns shows that the human element is still pivotal to the discussion. The dualism reinforces the multidisciplinary approach required in the study of data privacy, in which technological progress must be coupled with an understanding of human behavior to guarantee privacy-preserving solutions.

From the keyword co-occurrences and densities visualizations, we can conclude

that "data privacy" becomes the main concept of this domain, which is highly related to the latest technological innovations such as AI, deep learning, and big data analytics. These findings confirm a change in the research focus from classical methods and techniques of ensuring information security, such as cryptography or network security, to new, advanced ones necessary in the environment powered by AI. New keywords like "federated learning" and "privacy-preserving techniques" become quite popular due to the search for novel means to maintain a balance between data exploitation and privacy maintenance.

Furthermore, the visualized trend exhibits a clear evolution in time in terms of research topics, evolving from general investigations to more specific and future-oriented inquiries. Early works tended to concentrate on data protection and the creation of privacy guidelines, whereas modern investigations tend to discuss ethical issues, regulatory frameworks, and sustainable data management. Terms like "ethical technology" and "sustainable development" show that data privacy is not merely a technical matter but a socio-technical issue. This trend reflects the contemporary regulatory and social needs for data privacy research.

The present investigation will contribute to the existing body of knowledge by offering an exhaustive map of the research domain of data privacy between 2010 and 2026, emphasizing the changes that have taken place within the structure, intellectuality, and themes of this area of research. The results indicate that although considerable advancements have been achieved in terms of technicalities, there is still a gap that needs to be bridged in order to facilitate an integrated approach to data privacy issues.

## 4. CONCLUSION

This study presents an extensive bibliometric assessment of the development of data privacy literature from 2010 to 2026, illustrating the rapid growth and increasing complexity within the research domain. In terms of results, it can be observed that the area of data privacy research is becoming a highly interdisciplinary field, combining various technological innovations, such as AI and machine learning algorithms, with behavior-related, organizational, and regulatory studies. As for collaboration, there is an evident mixture of regional clustering with some signs of global convergence. As for citations, both technical and user-related literature have been identified as essential for forming the basis of research on the topic. Also, it is important to highlight the transition from security-based approaches to innovative privacy technologies and ethics-driven solutions, as indicated by the thematic assessment.

## REFERENCES

[1] A. S. Ali, Z. F. Zaaba, and M. M. Singh, "The rise of 'security and privacy': bibliometric analysis of computer privacy research," *Int. J. Inf. Secur.*, vol. 23, no. 2, pp. 863–885, 2024.

[2] S. Hossain, M. Amees, H. Zaffar, and N. Ahmad, "Measuring scholarly influence: a bibliometric analysis of data privacy on social media research," *Inf. Comput. Secur.*, vol. 34, no. 1, pp. 47–65, 2026.

[3] K. Kumar and M. Khari, "A privacy preserving mechanisms to secure data driven approaches in industrial internet of things: A bibliometric analysis," *Peer-to-Peer Netw. Appl.*, vol. 19, no. 3, p. 70, 2026.

[4] İ. H. Efendioğlu, "Marketing and Data Privacy: A Bibliometric Analysis," *Kocaeli Üniversitesi Sos. Bilim. Derg.*, vol. 2, no. 48, pp. 14–42, 2024.

[5] A. Valencia-Arias, J. D. González-Ruiz, L. Verde Flores, L. Vega-Mori, P. Rodríguez-Correa, and G. Sánchez Santos, "Machine learning and blockchain: A bibliometric study on security and privacy," *Information*, vol. 15, no. 1, p. 65, 2024.

[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE internet things J.*, vol. 3, no. 5, pp. 637–646, 2016.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*, Springer, 2006, pp. 265–284.

[8] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Commun. Soc.*, vol. 15, no. 5, pp. 662–679, 2012.

[9] F. Meyer *et al.*, "The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics*, vol. 9, no. 1, p. 386, 2008.

[10] D. J. Kim, D. L. Ferrin, and H. R. Rao, "A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents," *Decis. Support Syst.*, vol. 44, no. 2, pp. 544–564, 2008.

[11] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model," *Inf. Syst. Res.*, vol. 15, no. 4, pp. 336–355, 2004.

[12] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[13] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[14] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming–a review," *Agric. Syst.*, vol. 153, pp. 69–80, 2017.

[15] P. A. Pavlou, H. Liang, and Y. Xue, "Understanding and mitigating uncertainty in online exchange relationships: a principal–agent perspective1," *MIS Q.*, vol. 31, no. 1, pp. 105–136, 2007.