

AI-DRIVEN PERSONALIZATION AND PRIVACY ISSUES: A BIBLIOMETRIC MAPPING OF INTELLECTUAL, CONCEPTUAL, AND THEMATIC EVOLUTION

***Neha Sabu, **Dr. Vinod A S**

Abstract

This study conducts a bibliometric analysis of 500 documents recovered from the Dimensions database (2023-2026) to map the intellectual and thematic structure of AI-driven personalization and privacy study. Using Bibliometrix (R) and VOSviewer, publication trends, citation dynamics, collaboration networks, and thematic evolution are examined. Results uncover a compound annual expansion rate of 24.36 per cent, with China and the United States leading worldwide contributions. Thematic mapping identifies artificial intelligence-precision medicine-humans as motor themes, while demographic-driven personalization constitutes essential study streams. Emerging clusters suggest a shift toward wearable devices, deep learning, and big and substantial language models. Privacy management, nonetheless, remains comparatively underdeveloped relative to technological innovation. The study proposes a future inquiry agenda emphasising explainable AI, privacy-preserving architectures, and regulatory alignment frameworks.

Keywords:- : Artificial Intelligence, Personalization, Privacy, Bibliometric Analysis, Data Governance.

*A*I-driven personalization has become a primary and principal paradigm in digital transformation, allowing predictive medical care, adaptive recommendation systems, targeted advertising, and behaviour-based information delivery.

These systems depend heavily on wide-ranging and broad personal data extraction, algorithmic profiling, and behavioural inference, raising vital privacy concerns.

Regulatory frameworks such as the General Data Protection Regulation

**Neha Sabu, Research Scholar, Department of Commerce, Mahatma Gandhi College, Thiruvananthapuram, University of Kerala, E-mail: nehasabu075@gmail.com*

***Dr. Vinod A S, Associate Professor and Head, Department of Commerce, Mahatma Gandhi College, Thiruvananthapuram, University of Kerala, E-mail: vinodsreedher@gmail.com*

(GDPR) and the EU AI Act have intensified scrutiny over automated decision-making, making it necessary and fundamental to comprehend the structural evolution of this study domain.

The period 2023–2026 denotes a decisive expansion phase driven by the proliferation of deep learning and huge and massive language models (LLMs). Despite accelerating inquiry volume, the intellectual and thematic structure of this domain remains fragmented. This study addresses that gap through a methodical and organized bibliometric mapping of AI-driven personalization and privacy scholarship.

2. Scope of the Study

This study is confined to English-language peer-reviewed publications indexed in the Dimensions database from January 2023 to January 2026. It focuses exclusively on research at the intersection of AI-driven personalization and privacy, encompassing healthcare, digital platforms, and consumer ecosystems. The analysis does not include purely technical AI studies lacking a personalization or privacy dimension, editorials, or non-peer-reviewed documents. The scope is further bounded to bibliometric performance analysis, social structure mapping, conceptual clustering, and thematic evolution - without primary data collection or experimental validation.

3. Significance of the Study

This study offers several contributions. First, it offers the first systematic post-2023 bibliometric synthesis of AI personalization and privacy inquiry, capturing the generative

AI expansion era. Second, it discloses a structural asymmetry between fast technological innovation and comparatively underdeveloped privacy management scholarship; a disparity with direct regulatory outcomes. Third, it maps worldwide collaboration networks, identifying local inquiry silos and fragmented interdisciplinary integration. Fourth, it equips policymakers, researchers, and system designers with an organized knowledge foundation to prioritise explainable AI, federated learning, and regulatory adherence architectures. Fifth, it proposes an actionable study agenda to bridge the management gap before autonomous personalization systems outpace regulatory capability.

4. Objectives of the Study

The primary objective is to systematically map the intellectual, social, and conceptual structure of research on AI-driven personalization and privacy during 2023–2026. Specifically, the study aims to:

- Analyse publication and citation dynamics to assess growth trajectory and scholarly impact.
- Identify productive authors, influential journals, and leading countries to map global knowledge production.
- Explore conceptual structure through keyword co-occurrence and thematic mapping.
- Assess the balance between technological innovation and privacy governance within AI personalization scholarship.

5. Review of Literature

AI-driven personalization leverages large-scale data collection, machine learning, and behavioural profiling to tailor services to individual users (Goodfellow et al., 2016). In medical care, it has advanced precision medicine by integrating genomic, behavioural, and clinical data for individualized treatment (Rieke et al., 2020). Deep learning architectures have significantly enhanced personalization accuracy in recommender systems and predictive condition analytics, while wearable devices have expanded real-time adaptive abilities.

Federated learning has emerged as a promising privacy-preserving approach, enabling decentralized model training without centralized data storage (Kairouz et al., 2021). Despite this, privacy vulnerabilities such as membership inference and model inversion attacks demonstrate persistent data leakage risks (Shokri et al., 2017). Differential privacy frameworks introduce statistical noise to datasets to mitigate these risks, though at a performance cost (Dwork, 2008).

Ethical frameworks emphasise transparency, accountability, fairness, and human oversight (Floridi & Cowls, 2019; Mittelstadt et al., 2016). GDPR requires explainability and informed consent for automated profiling, yet rapid technological evolution creates persistent governance gaps (Voigt & Von dem Bussche, 2017). Existing research remains fragmented across technical AI innovation, biomedical personalization, and privacy governance domains, creating a gap that systematic bibliometric mapping can address.

6. Methodology

6.1 Data Source and Search Strategy

The dataset was retrieved from the Dimensions database in January 2026. A structured Boolean search string combining terms related to artificial intelligence, personalization, and privacy was applied to titles, abstracts, and keywords, restricted to English-language documents published between 2023 and 2026. The initial search returned 538 records; after removing 18 duplicates and 20 incomplete or non-English records, 500 documents were retained.

6.2 PRISMA Screening Summary

Records identified: 538 | Duplicates removed: 18 | Non-English/incomplete excluded: 20 | Final dataset: 500 documents.

6.3 Inclusion and Exclusion Criteria

Included: peer-reviewed journal articles, book chapters, and conference proceedings; published 2023–2026; focus on AI-driven personalization with privacy implications; English language.

Excluded: purely technical AI studies without personalization context; non-peer-reviewed reports; editorials or commentaries; duplicate records.

6.4 Analytical Tools and Hypotheses

Bibliometrix (R) was used for performance and thematic analysis; VOS viewer for co-authorship and keyword co-occurrence network visualization. The analysis was structured into: (i) performance analysis, (ii) social structure mapping, and (iii) conceptual structure analysis.

7. Results and Interpretation

7.1 Performance Analysis

The dataset comprises 500 documents drawn from 267 sources, involving 2,056 authors. The compound annual increase rate (CAGR) is 24.36 per cent, indicating accelerated expansion uniform and unchanging with H1. The standard and normal citations per document stand at 10.94, indicating fast and rapid knowledge diffusion. The relatively young usual and customary document age (1.19 years) verifies that the area remains in its first expansion stage. High collaboration intensity (4.51 co-authors per document) signals interdisciplinary engagement across computer science, medical care, and social sciences.

7.2 Annual Scientific Production

Publication output rose from 13 documents in 2023 to 370 in 2025,

representing exponential growth of 302.2 per cent between 2024 and 2025 alone. This surge corresponds with widespread generative AI deployment and heightened global privacy debates, confirming H1. The 2026 figure reflects partial-year data and not declining interest (Table 1).

7.3 Citation Analysis

Earlier publications (2023–2024) (Table 2) exhibit substantially higher mean citation rates compared to 2025–2026 outputs, consistent with recency bias. The 2023 cohort averages 38.15 citations per article, nearly seven times the 2025 average of 5.83. This pattern demonstrates fast citation turnover and strong visibility for foundational contributions. Notably, influential work integrates AI performance with ethical or regulatory considerations rather than purely algorithmic optimization—a finding relevant to H2.

Table 1
Annual Scientific Production (2023–2026)

Year	Articles Published	Cumulative Total	Growth Rate (%)
2023	13	13	—
2024	92	105	+607.7
2025	370	475	+302.2
2026 (partial)	25	500	—
Total	500	—	24.36% CAGR

Source: Dimensions database; Bibliometrix (R).

Table 2
Citation Analysis by Year

Year	N	Mean TC/Article	Mean TC/Year	Citable Years
2023	13	38.15	9.54	4
2024	92	30.61	10.20	3
2025	370	5.83	2.92	2
2026	25	0.00	0.00	1

Source: Dimensions database; Bibliometrix (R).

7.4 Most Relevant Sources

Dominance of applied biomedical outlets—Sensors (27), Scientific Reports (18), Healthcare (16), and Cureus (15) - confirms that AI-driven personalization research is primarily embedded in biomedical and technological contexts, reflecting precision medicine as the central application domain (Table 3). The complete absence of law or governance-specialist journals in the top ten provides strong evidence for H4: privacy scholarship remains peripheral to technological innovation.

7.5 Most Productive Authors

Zhang J leads with seven articles, followed by Giansanti D (6), Wang H (6), and Zhang L (6) (Table 4). The fractionalized article counts reveal that many contributions are collaborative, with Giansanti D holding the highest fractionalized score (3.25), indicating relatively independent authorship compared to others. The dominance of Chinese and European-affiliated authors reflects the country-level production pattern discussed below.

Table 3
Top 10 Most Productive Journals

Rank	Journal / Source	Articles
1	Sensors	27
2	Scientific Reports	18
3	Healthcare	16
4	Cureus	15
5	Frontiers in Artificial Intelligence	12
6	Frontiers in Digital Health	11
7	Digital Health	10
8	Acta Psychologica	9
9	Frontiers in Public Health	8
10	Frontiers in Psychology	7

Source: Dimensions database; Bibliometrix (R).

Table 4
Most Productive Authors

Author	Articles	Fractionalized Articles
Zhang J	7	1.73
Giansanti D	6	3.25
Wang H	6	2.33
Zhang L	6	1.24
Calabra RS	5	0.99
Chen Y	5	0.72
Liu X	5	1.26
Olawade DB	5	1.10
Wang Y	5	0.98

Source: Dimensions database; Bibliometrix (R).

7.6 Country-Level Analysis

China leads with 81 articles (16.2 per cent), followed by the USA (61; 12.2 per cent), Italy (27; 5.4 per cent), and India (23; 4.6 per cent) (Table 5). The USA’s MCP rate of 0.0 per cent indicates that all 61 US publications are domestically authored, suggesting inward-looking research activity. Australia’s MCP rate of 75.0 per cent and Canada’s of 41.2 per cent indicate high international collaboration intensity despite lower overall volume. The overall international co-authorship rate of 22.2 per cent confirms H3: the majority of research is produced within national silos, limiting global knowledge integration.

7.7 Social Structure: Co-Authorship Network

The co-authorship network comprises 16 authors organized into three clusters connected by 54 collaboration links (total link strength: 64). The largest cluster (Olawade, Egbon, Boussios, Mohammed and associates) represents healthcare-oriented personalization research with tight intra-group ties. The second cluster (Adeyeye, Weerasinghe, Teke, Clement

David-Olawade) reflects thematic specialisation in wearable technologies. The third, smaller cluster bridges the two dominant groups but with sparse inter-cluster connectivity - indicating that the field evolves through parallel rather than fully integrated research streams. This fragmentation corroborates H3 at the author level.

7.8 Conceptual Structure: Keyword Analysis

Keyword co-occurrence analysis positions ‘humans,’ ‘artificial intelligence,’ and ‘precision medicine’ as dominant concepts, confirming biomedical primacy in personalization discourse. Terms such as ‘GDPR,’ ‘data protection law,’ and ‘regulation’ have negligible co-occurrence frequency, reinforcing H4. Machine learning, wearable devices, and deep learning appear as secondary but growing clusters, consistent with the emerging theme classification below.

7.9 Thematic Map

The strategic diagram confirms H2: motor themes dominated by artificial intelligence, precision medicine, and

Table 5
Country-Level Production and Collaboration

Country	Articles	Articles %	SCP	MCP	MCP %	Collaboration
China	81	16.2	59	22	27.2	High
USA	61	12.2	61	0	0.0	Domestic
Italy	27	5.4	23	4	14.8	Moderate
India	23	4.6	19	4	17.4	Moderate
Canada	17	3.4	10	7	41.2	High
Saudi Arabia	15	3.0	15	0	0.0	Domestic
Australia	12	2.4	3	9	75.0	Very High
Germany	12	2.4	8	4	33.3	High

Source: Dimensions database; Bibliometrix (R). SCP = Single Country Publications; MCP = Multiple Country Publications.

Table 6
Thematic Map Classification (Strategic Diagram)

Theme Quadrant	Keywords / Clusters	Interpretation
Motor Themes (High C, High D)	AI, Precision Medicine, Humans	Core, well-developed, structurally central to field
Niche Themes (Low C, High D)	Large Language Models, Language, Communication	Specialised but not yet mainstream
Emerging Themes (High C, Low D)	Algorithms, Wearable Devices, Deep Learning	Growing relevance; transitioning toward real-time AI
Basic/Foundational (Low C, Low D)	Female, Male, Adult	Demographic variables; foundational personalization units

Source: Dimensions database; VOSviewer. C = Centrality; D = Density.

humans are both central and well-developed. Demographic variables (female, male, adult) constitute foundational but non-innovative personalization categories. Wearable devices, deep learning, and algorithms are positioned as emerging themes, signalling a transition toward real-time adaptive personalization. Large language models appear as niche themes - specialised but not yet structurally integrated—reflecting the nascent stage of generative AI personalization research.

8. Discussion

The analysis reveals a structural imbalance: technological personalization innovation consistently outpaces privacy governance integration. The field is innovation-driven rather than regulation-driven. As AI systems become increasingly autonomous - particularly with the emergence of LLM-based personalization - this imbalance introduces risks of inference attacks, synthetic profiling, and unconstrained behavioural prediction. The absence of law and governance journals from the top sources, combined with the peripheral

positioning of GDPR and regulatory keywords, indicates that privacy is currently treated as an external constraint rather than a design principle. The geographic concentration of output in China and the USA, alongside low international co-authorship rates, signals fragmented global knowledge exchange. Regulatory discourse is especially under-represented in high-output countries, which may reflect divergent national approaches to AI governance. Bridging this gap requires deliberate interdisciplinary engagement between computer scientists, legal scholars, and ethicists.

9. Findings and Suggestions

9.1 Key Findings

- The field recorded a CAGR of 24.36 per cent, growing from 13 publications in 2023 to 370 in 2025—driven by generative AI proliferation.
- China (16.2 per cent) and the USA (12.2 per cent) dominate global output, but the USA operates almost exclusively within domestic co-authorship networks (MCP = 0 per cent).

- Precision medicine and clinical AI personalization constitute the intellectual core of the field, while commercial and digital platform personalization is under-represented.
- Privacy governance terms (GDPR, regulation, data protection law) are marginal in keyword co-occurrence networks, confirming a critical governance gap.
- Wearable devices, deep learning, and algorithms are transitioning from emerging to mainstream themes, signalling a shift toward real-time personalization ecosystems.
- Large language models appear as niche themes, indicating an impending transformation wave not yet structurally integrated into the field.
- Co-authorship fragmentation into parallel clusters limits interdisciplinary knowledge synthesis across computer science, law, and social sciences.

9.2 Suggestions and Recommendations

- Privacy-by-design integration: Future personalization system architectures must embed privacy preservation at the design stage rather than as a post-hoc constraint. Federated learning and differential privacy should be adopted as standard components.
- Explainable AI mandates: Regulatory frameworks should require explainability certifications for AI personalization systems deployed in healthcare, finance, and education -

domains with high sensitivity to algorithmic bias.

- Cross-disciplinary research funding: Funding agencies should incentivise joint research teams comprising AI engineers, legal scholars, ethicists, and behavioral scientists to address the identified governance–technology gap.
- LLM-specific privacy frameworks: As LLM-based personalization grows, dedicated governance instruments addressing inference risks, synthetic profiling, and data minimization in generative models must be developed.
- Global governance harmonization: International regulatory alignment between GDPR (EU), DPDP Act (India), and emerging frameworks in China and the USA is necessary to prevent regulatory arbitrage in AI personalization.
- Longitudinal bibliometric studies: Future research should extend the temporal window beyond 2026 and incorporate Scopus and Web of Science to validate growth trajectories and thematic evolution across indexing platforms.

10. Limitations

This study relies exclusively on the Dimensions database; inclusion of Scopus or Web of Science may alter citation counts and journal rankings. The 2023–2026 timeframe captures the generative AI expansion phase but omits foundational pre-2023 literature, limiting longitudinal perspective. Recency bias in citation metrics may underestimate the

long-term impact of recent publications. The co-authorship network is limited to the 500-document sample and may not fully represent the global collaboration landscape.

11. Conclusion

This bibliometric study demonstrates that AI-driven personalization and privacy research is in an accelerated expansion phase, particularly within healthcare and precision medicine domains. Motor themes centred on artificial intelligence and precision medicine dominate the intellectual core, while privacy regulation and governance remain peripheral. Emerging themes - wearable technologies, algorithmic personalization, and large language models - signal a transition toward real-time, adaptive, and

autonomous personalization ecosystems that existing governance frameworks are ill-equipped to address.

The study confirms four hypotheses: significant growth (H01), technological dominance over governance in thematic structure (H02), limited international collaboration (H03), and peripheral positioning of privacy governance (H04). The findings collectively call for integrating privacy-preserving architectures, explainable AI mechanisms, and regulatory compliance models into system design as first-order requirements rather than afterthoughts. Bridging the gap between technological performance optimization and normative governance frameworks is essential for sustainable, rights-respecting AI deployment.

References

1. Bansal, G., Zabedi, F. M., & Gefen, D. (2016). *The impact of personal dispositions on information privacy concerns*. *Information Systems Research*, 27(3), 583-603.
2. Dwork, C. (2008). *Differential privacy: A survey of results*. *Lecture Notes in Computer Science*, 4978, 1-19.
3. Floridi, L., & Cowls, J. (2019). *A unified framework of five principles for AI in society*. *Harvard Data Science Review*, 1(1).
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
5. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., et al. (2021). *Advances and open problems in federated learning*. *Foundations and Trends in Machine Learning*, 14(1-2), 1-210.
6. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). *The ethics of algorithms: Mapping the debate*. *Big Data & Society*, 3(2), 1-21.
7. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., et al. (2020). *The future of digital health with federated learning*. *NPJ Digital Medicine*, 3(1), 119.
8. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). *Membership inference attacks against machine learning models*. *IEEE Symposium on Security and Privacy*, 3-18.
9. Voigt, P., & Von dem Bussche, A. (2017). *The GDPR: A Practical Guide*. Springer.
10. European Parliament. (2024). *EU AI Act: Regulation on Artificial Intelligence*. *Official Journal of the European Union*.
11. Aria, M., & Cuccurullo, C. (2017). *Bibliometrix: An R-tool for comprehensive science mapping analysis*. *Journal of Informetrics*, 11(4), 959-975.
12. Van Eck, N. J., & Waltman, L. (2010). *Software survey: VOSviewer, a computer program for bibliometric mapping*. *Scientometrics*, 84(2), 523-538.