

AI-Driven Personalization and Privacy Issues: A Study of The Literature Using Citation Network Analysis

Ms Neha Sabu

Research Scholar,

Department of Commerce,

Mahatma Gandhi College, Thiruvananthapuram,

University of Kerala

Email id: nehasabu075@gmail.com

Dr Vinod A.S

Associate Professor and Head,

Department of Commerce,

Mahatma Gandhi College,

Thiruvananthapuram, University of Kerala

Email id: vinodsreedher@gmail.com

ABSTRACT:

AI-based personalisation has become one of the key components of current technological solutions. Personalisation technologies allow for adjusting the content, services and recommendations according to the preferences of individual users. Although the collection and inference of personal information is extremely beneficial, it creates severe privacy concerns. The present study offers a citation network analysis (CNA) of 362 primary articles on AI-based personalisation and privacy, which are collected from Scopus and Web of Science sources and result in a citation network of 28,967 nodes and 31,114 directed links between 2013 and 2026. For the purposes of main path analysis, we employed Pajek software, while community detection was conducted using Gephi. Three main topics were identified, namely (1) AI-based marketing personalisation and consumer behaviour, (2) human-computer interaction, explainability and user trust, and (3) ethics, regulation and digital well-being. Four bridge nodes exist on the central route connecting the clusters. The node with the highest betweenness centrality score is 250. Kumar et al., (2024), on the other hand, is the most cited document in the corpus with 10 local citations. Based on the findings, the area can be considered only partly siloed since there seems to be a lack of integration between technical privacy solutions and regulatory discussion.

Keywords: AI personalization, privacy, citation network analysis, recommender systems, user trust, data privacy, GDPR, algorithmic transparency, digital well-being, main path analysis, bibliometrics, explainable AI

INTRODUCTION

Artificial intelligence-driven personalization has essentially transformed how digital platforms engage with users. Recommendation engines, directed advertising systems, and adaptive interfaces now underpin the business models of social media, e-commerce, medical care, and news distribution globally. These systems generate substantial economic and experiential value, but they concurrently generate deep tensions with individual privacy rights, autonomy, and notified permission.

Privacy concerns in AI personalization function across numerous and manifold dimensions. At the data collection level, personalization demands continuous behavioral monitoring. At the inference level, AI models derive sensitive personal attributes; political principles, condition conditions, emotional states, and financial vulnerability; from ostensibly innocuous behavioral data. At the management level, the asymmetry between platforms and users has drawn regulatory attention internationally, producing landmark frameworks such as the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Despite substantial academic activity spanning computer science, marketing, information systems, law, and psychology, the intellectual structure of this domain remains unmapped. Which papers are essential? How do dissimilar study traditions relate to one another? How has the discourse evolved over time? Citation network analysis (CNA) addresses these queries by treating publications as nodes and citations as guided edges, thereby revealing the prominence of individual texts, the clustering of intellectual groups, and the historical trajectory of an area (McLaren & Bruner, 2022).

The present study applies CNA to a collection of 362 primary publications, forming a network of 28,967 nodes and 31,114 citation connections. Main path analysis in Pajek traces the primary and principal intellectual backbone, and Gephi is used for community detection and visualization. This paper pursues three study queries: (RQ1) What is the overall structure of the citation network and which publications are most notable and distinguished? (RQ2) What distinct intellectual clusters exist and what are their defining themes? (RQ3) What does the principal path disclose about bridge nodes joining study traditions?

METHODOLOGY

A methodical and organized search of Scopus and Web of Science was conducted utilizing a three-block search string unifying AI personalization conditions ("AI personalization", "algorithmic personalization", "recommender system *"), privacy conditions ("privacy", "data privacy", "user privacy", "personal data"), and AI methodology

The Voice of Creative Research

Vol. 8 & Issue 2 (April 2026)

conditions (“machine learning”, “deep learning”, “neural network *”). The search was restricted to peer-reviewed journal articles and conference proceedings in English. No publication date constraint was applied. After deduplication, 362 primary publications were kept (2013–2026).

Full bibliographic records; including full and comprehensive reference lists; were exported and used as input for network construction in Pajek (version 5.06). Each publication was represented as a node and each citation relationship as a guided arc, generating the full and comprehensive network of 28,967 nodes and 31,114 citation connections. Main path analysis was performed utilizing the Search Path Link Count (SPLC) traversal weight strategy, which identifies the citation chain most responsible for transmitting intellectual influence from fundamental origin papers to contemporary and ongoing literature.

Community detection on the primary paper subnetwork was conducted in Gephi (version 0.9.7) applying the Louvain modularity algorithm at resolution parameter 1.0. Network metrics; indegree significance (local citation count), betweenness prominence, and density; were calculated for all nodes. Qualitative analysis of the three biggest clusters followed the approach of Kusumastuti et al. (2016), identifying the pioneering publication, the highest-cited publication, and the most latest and current highly-cited contribution within each cluster.

RESULTS

3.1 Network Overview

The full and comprehensive citation network comprised 28,967 publications and 31,114 managed citation connections. The primary collection of 362 papers spanned 2013 to 2026, with a pronounced concentration in latest years: 249 papers (68.8%) released in 2025 and 42 in 2024. This reflects the quick and fast acceleration of study following the mainstream deployment of big and substantial language models from 2022 onwards. The most represented journals were IEEE Access (n=17), Electronics; MDPI (n=11), Applied Sciences; MDPI (n=11), Sustainability; MDPI (n=10), and the International Journal of Human-Computer Interaction (n=7), indicating the field’s interdisciplinary character.

Table 1 presents the ten most notable and distinguished primary papers ranked by local citation count; the number of citations received from other papers within the collection. Kumar et al. (2024, International Journal of Information Management) and Kumar et al. (2019, California Management Review) appear as the utmost and preeminent internally-cited papers (LC=10 and LC=8 respectively), with Sundar and Kim (2020, Journal of Computer-Mediated Communication) achieving the supreme and paramount worldwide citation count (GC=481).

Table 1: Top 10 Most Prominent Primary Papers by Local Citation Count

Rank	First Author / Year	Year	LC	GC	Journal
1	Kumar V et al.	2024	10	157	Int J Inform Manage
2	Kumar V et al.	2019	8	337	Calif Manage Rev

The Voice of Creative Research

Vol. 8 & Issue 2 (April 2026)

Rank	First Author / Year	Year	LC	GC	Journal
3	Ahmed S et al.	2025	5	36	Int J Hum-Comput Int
4	Sundar SS & Kim J	2020	5	481	J Comput-Mediat Comm
5	Teepapal T	2025	5	50	Comput Hum Behav
6	Gupta S et al.	2020	3	118	J Interact Mark
7	Boucher EM et al.	2021	3	213	Expert Rev Med Devic
8	Mende M et al.	2024	3	43	J Serv Res-US
9	Kunz WH et al.	2024	3	85	J Res Interact Mark
10	Lim JS et al.	2022	2	94	Technology in Society

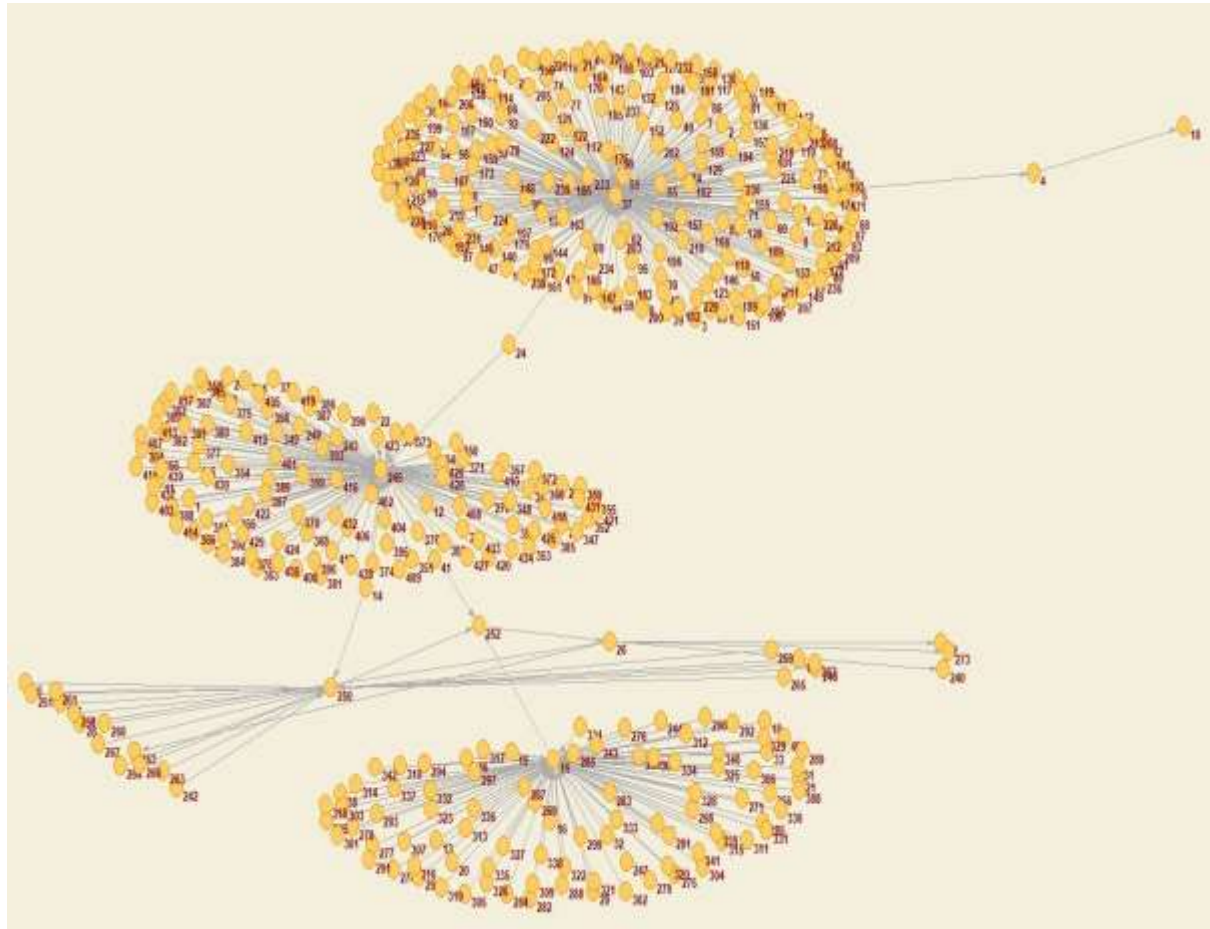
Note: LC = Local Citation Count (within corpus). GC = Global Citation Count (WoS).

3.2 Main Path Analysis

Main path analysis revealed a multi-hub topology comprising three huge and massive elliptical clusters linked by a sequential chain of bridge nodes (see Figure 1). The three clusters correspond to the three primary intellectual traditions identified by community detection. Cluster A (upper ellipse, Pajek hub node 37, ~240 nodes) is the oldest and densest cluster, rooted in AI personalization study in marketing and consumer behavior. Cluster B (middle-left ellipse, hub node 245, ~180 nodes) denotes the human-computer interaction and AI explainability wave. Cluster C (reduced ellipse, hub node 15, ~150 nodes) is the most recently constituted cluster, tackling ethics, regulation, and digital well-being.

Four bridge nodes connect these clusters on the primary path. Bridge node 24 connects Cluster A to Cluster B and denotes the pivot from business-oriented personalization discourse toward algorithmic transparency. Bridge node 14 serves as a secondary connector along the primary path spine. Bridge node 250; the highest-betweenness node in the total and full network; connects all three clusters and serves as the vital pivot from technical personalization discourse toward privacy, ethics, and user well-being concerns. The exit chain 252→26 connects Cluster C to a set of peripheral nodes (IDs: 259, 265, 273, 240) representing specialized sub-topics in neighbouring fields. Nodes 4 (Behera et al., 2020, Journal of Retailing and Consumer Services) and 10 (Rauh et al., 2020, Journal of Public Economics) appear as isolated peripheral nodes connected to Cluster A by single citation links, indicating early citations of AI personalization from adjacent disciplines.

Figure 1: Main Path Analysis of the AI Personalization and Privacy Citation Network



(WoS, Pajek)

3.3 Research Clusters: Qualitative Analysis

Cluster A; AI-Driven Marketing Personalization: This cluster is rooted in marketing science and consumer behaviour. Its pioneering contribution is Kumar et al. (2019, *California Management Review*), which founded a framework for AI's role in customer engagement and firm value, and remains the second-most internally cited paper in the collection (LC=8, GC=337). The highest-cited paper overall is Sundar and Kim (2020, GC=481), which introduced the chief model explaining how AI personalization influences impressions of origin credibility and user agency. The present and existing anchor of the cluster is Kumar et al. (2024, LC=10), indicating how the marketing tradition has absorbed generative AI concerns. Privacy enters Cluster A primarily through concerns about monitoring capitalism, behavioural targeting, and the commodification of personal data.

Cluster B; Human-Computer Interaction and User Trust: Cluster B emerges from the HCI and information systems traditions. Its intellectual center is Conati et al. (2021, *Artificial Intelligence*, GC=118), which examined how explainable AI (XAI) techniques can make

The Voice of Creative Research

Vol. 8 & Issue 2 (April 2026)

algorithmic decisions legible to users in tailored adaptive systems. Lim et al. (2022, *Technology in Society*, GC=94) and Shin (2021, *Journalism Studies*) anchored empirical study on how AI monitoring views influence trust and data-sharing willingness. Bridge node 24 joining Cluster A to Cluster B aligns to the conceptual pivot from solely output-focused personalization inquiry to process-focused study asking how users understand and control algorithmic curation.

Cluster C; Ethics, Regulation, and Digital Well-being: Cluster C are the most recently formed cluster, driven by GDPR imposition (2018), the deployment of conversational AI (2022+), and growing public concern about algorithmic damage. Kozyreva et al. (2021, *Humanities and Social Sciences Communications*, GC=78) supplied the intellectual foundation by arguing that personalization undermines epistemic autonomy. Teepapal (2025, *Computers in Human Behavior*, LC=5) is the most new highly-cited contribution, empirically examining how AI condition recommendations influence trust, recognized privacy risk, and behavioral aim. A notable sub-cluster addresses healthcare AI personalization (Boucher et al., 2021; Abbasian et al., 2024), where privacy stakes are supreme and paramount.

DISCUSSION

The citation network discloses a domain that is concurrently fragmented and converging. Three intellectually distinct clusters; rooted in marketing science, HCI, and ethics respectively; have developed largely in parallel before being linked by bridge publications of unusually high betweenness prominence. The dominance of bridge node 250 as the network's central connector implies that the transition from technical personalization discourse to privacy and ethics concerns was concentrated and critical and crucial alternatively than incremental, steady and regular with paradigm-shift dynamics recorded in other interdisciplinary fields (McLaren & Bruner, 2022).

The sustained prominence of Kumar et al. (2019; 2024) as the two supreme and paramount locally-cited papers verifies that the marketing science tradition remains the intellectual center of gravity of the area. This is a notable discovery for researchers in computer science or law seeking field-level impact: engaging with marketing frameworks appears to be an effective approach for achieving broad cross-cluster citation. Conversely, the relative separation of the technical privacy approaches sub-literature; particularly differential privacy and federated learning approaches; from the regulatory and moral Cluster C denotes an essential gap. Researchers in Cluster C seldom cite the mathematical privacy literature, and vice versa.

The exceptional concentration of publications in 2024–2026 (n=316 of 362, or 87.3%) signals a domain in swift expansion. The recency effect; whereby latest papers have had inadequate time to gather citations; means that contemporary and ongoing local citation counts understate the future influence of several 2025 publications. Longitudinal recurrence of this CNA in three to five years will be essential to fully characterize the field's developing structure, especially as big and substantial language model-based personalization generates new study clusters not yet visible in this snapshot.

CONCLUSION

This citation network analysis offers the first methodical map of the intellectual structure of the AI-driven personalization and privacy literature. Analyzing 28,967 publications and 31,114 citation connections, primary path analysis identified three main clusters; AI marketing personalization, HCI and user trust, and ethics/regulation/well-being; connected by four high-betweenness bridge nodes. The most notable and distinguished papers are Kumar et al. (2024; 2019), Sundar and Kim (2020), and Boucher et al. (2021), indicating the field's dual anchoring in marketing science and applied digital condition.

The key finding is that interdisciplinary integration depends on a small number of bridging publications rather than broad cross-cluster citation. Strengthening bridges between technical privacy solutions and regulatory/ethical discourse is the most urgent intellectual need in this field. Future research should apply longitudinal CNA, co-word keyword analysis, and author collaboration network analysis to track how the structure evolves as AI personalization continues its rapid expansion into healthcare, education, and public services.

Acknowledgement

The authors acknowledge that this is an original and unpublished work. The research data were retrieved from Scopus and Web of Science databases. Network analysis was performed using Pajek (version 5.06) and Gephi (version 0.9.7). No funding was received for this study. The authors declare no conflict of interest.

The author acknowledges the application of artificial intelligence instruments, specifically ChatGPT, to help in enhancing the clearness, structure, and language of this manuscript. The intellectual material, analysis, and interpretations presented in this study continue the author's initial work.

REFERENCES

- Ahmed, S., et al. (2025). AI-driven personalization and privacy in human-computer interaction. *International Journal of Human-Computer Interaction*.
<https://doi.org/10.1080/10447318.2024.2400395>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1).
- Behera, R. K., et al. (2020). Co-experience in customer journey: A matter of trust on e-retail platforms. *Journal of Retailing and Consumer Services*, 53, 101993.
<https://doi.org/10.1016/j.jretconser.2019.03.026>
- Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., & Zilca, R. (2021). Artificially intelligent coaching to improve mental health: Pilot randomized controlled trial. *Expert Review of Medical Devices*, 18(3-4), 347-360.
<https://doi.org/10.1080/17434440.2021.2013200>

- Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298, 103503. <https://doi.org/10.1016/j.artint.2021.103503>
- De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge University Press.
- Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference (TCC)* (pp. 265–284).
- European Parliament and Council. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Official Journal of the European Union, L119, 1–88.
- Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T., & Potapov, D. (2020). Digital analytics: Modeling for insights and new methods. *Journal of Interactive Marketing*, 51, 26–43. <https://doi.org/10.1016/j.intmar.2020.04.003>
- Kang, H., et al. (2022). Algorithmic news personalization and communicative autonomy. *Journal of Computer-Mediated Communication*, 27, zmac014. <https://doi.org/10.1093/jcmc/zmac014>
- Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2021). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Humanities and Social Sciences Communications*, 8, 87. <https://doi.org/10.1057/s41599-021-00787-w>
- Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the role of artificial intelligence in personalized engagement marketing. *California Management Review*, 61(4), 135–155. <https://doi.org/10.1177/0008125619859317>
- Kumar, V., et al. (2024). How AI is transforming personalization in information management. *International Journal of Information Management*, 77, 102783. <https://doi.org/10.1016/j.ijinfomgt.2024.102783>
- Kunz, W. H., et al. (2024). AI in service: A review and research agenda. *Journal of Research in Interactive Marketing*, 18(3), 1–31. <https://doi.org/10.1108/JRIM-06-2023-0176>
- Kusumastuti, S., Derks, M. G. M., Tellier, S., Di Nucci, E., Lund, R., Mortensen, E. L., & Westendorp, R. G. J. (2016). Successful ageing: A study of the literature using citation network analysis. *Maturitas*, 93, 4–12. <https://doi.org/10.1016/j.maturitas.2016.04.010>
- Lim, J. S., et al. (2022). The effect of AI surveillance on workers and organizations. *Technology in Society*, 69, 101965. <https://doi.org/10.1016/j.techsoc.2022.101965>
- McLaren, C. D., & Bruner, M. W. (2022). Citation network analysis. *International Review of Sport and Exercise Psychology*, 15(1), 179–198. <https://doi.org/10.1080/1750984X.2021.1989705>
- Mende, M., et al. (2024). Alexa, are you listening? AI-driven voice assistants and customer privacy. *Journal of Service Research*, 27(2), 1–28. <https://doi.org/10.1177/10946705231188676>
- Rauh, J. D., et al. (2020). Algorithmic targeting of social policy. *Journal of Public Economics*, 188, 104211. <https://doi.org/10.1016/j.jpubeco.2020.104211>