Privacy concerns in tourism: a systematic literature review using machine learning approach and bibliometric analysis

Hitesh Sharma, Praveen Ranjan Srivastava, Sajjad M. Jasimuddin, Zuopeng Justin Zhang and Ikram Jebabli

Abstract
Purpose — This study aims to provide a comprehensive analysis of the current privacy concerns in the tourism industry by uncovering the key factors leading to such concerns (i.e. smart public services, cyber security issues, consumer behaviour and governance). Using papers from multiple sources, the relationship between technology advancements and tourist’s privacy concerns has been established.

Design/methodology/approach — This study adopted a machine learning-based systematic literature review approach to find out the clusters. The study analysed 68 papers using the topic modelling approach. A four-cluster solution was considered to be most representative of the extant literature identified using bibliographic coupling. Finally, content analysis of the selected literature has been performed.

Findings — This study identified four factors majorly leading to privacy concerns amid increasing technological advancements. Moreover, these factors were found to have a dyadic relationship with technological advancements. To everyone’s amazement, sustainable tourism was also found to have led to privacy concerns among tourists along with a lack of governance and cyber security issues. Furthermore, cluster-wise future research directions are provided based on the content analysis.

Originality/value — This study contributes to the literature by systematically reviewing and identifying the four dimensions leading to privacy concerns. To the best of the authors’ knowledge, the study done is the only attempt to synthesize the extant literature on tourists’ privacy concerns using an unbiased scientific approach.

Keywords Privacy concerns, Tourism, Technological advancements, Topic modelling, Bibliometric analysis, Clustering

Paper type Literature review
Preocupación por la privacidad en el turismo: Una revisión sistemática de la literatura utilizando el enfoque de aprendizaje automático y el análisis bibliométrico.

Resumen

Finalidad/Objetivo: Este estudio proporciona un análisis exhaustivo de la actual preocupación por la privacidad en el sector turístico, poniendo al descubierto los factores clave que la generan (los servicios públicos inteligentes, los problemas de ciberseguridad, el comportamiento de los consumidores y la gobernanza). Gracias al uso de artículos de múltiples fuentes, se ha establecido la relación entre los avances tecnológicos y la preocupación de los turistas por la privacidad.

Diseño/Metodología/Enfoque: Este estudio adoptó un enfoque de revisión sistemática de la literatura basado en el aprendizaje automático para descubrir los conglomerados. El estudio analizó sesenta y ocho artículos utilizando el enfoque de modelización de temas. Se consideró que una solución de cuatro conglomerados era la más representativa de la literatura existente identificada mediante el acoplamiento bibliográfico. Por último, se realizó un análisis de contenido de la bibliografía seleccionada.

Hallazgos: En este estudio se identificaron cuatro factores principales que suscitan inquietud por la privacidad en medio de los crecientes avances tecnológicos. Además, se descubrió que estos factores tienen una relación diédrica con dichos avances. Para sorpresa de todos, se halló que el turismo sostenible, junto con la falta de gobernanza y los problemas de ciberseguridad, también suscitan entre los turistas preocupaciones por su privacidad. Finalmente, el análisis de contenido ofrece orientaciones para futuras investigaciones.

Originalidad: Este estudio contribuye a la literatura haciendo una revisión sistemática e identificando las cuatro dimensiones que conducen a la preocupación por la privacidad. Este estudio es el único intento de sintetizar la bibliografía existente sobre la preocupación de los turistas por su privacidad utilizando un enfoque científico imparcial.

Palabras clave Preocupación por la privacidad, Turismo, Avances tecnológicos, Modelización temática, Análisis bibliométrico, Conglomerado

Tipo de papel Revisión de literatura

1. Introduction

With the increased technology adoption to enhance travel, tourism is becoming more intelligent exponentially. However, the vast data that destinations and service providers have access to has raised worries about the privacy of visitor information (Afolabi et al., 2020). This severe tourism issue may impact the traveller’s behavioural intention. Using personalized services with technology is essential to keep up with the present and future advancements in the tourism industry (Kontogianni and Alepis, 2020). Mobile technology users are growing, and service providers and stakeholders in tourism are concentrating on creating cutting-edge applications to fulfil the needs of tourists (Chang and Shen, 2018). Thus giving rise to tourists’ privacy concerns (PCs). Despite the ubiquitous effect of PCs, tourism literature has paid less attention to it (Gong and Schroeder, 2022; Hall and Ram, 2020; Tussyadiah et al., 2019).

Travellers may now take images and share them with their friends, obtain knowledge to help them make better selections, and locate the closest points of interest thanks to technology (Buhalis and Foerste, 2015). They now use portable devices like they once did their desktop and laptop computers (Ozturk et al., 2017).

There is no widely agreed-upon definition of privacy. PCs are growing, especially as personal data is commercialized and used for power (Hall and Ram, 2020). Due to the transient nature of users, the duty of care owed to tourism business participants, and the many information and communication technology (ICT) jurisdictions pertaining to tourist moving nature, data storage and unique service encounters, privacy in the tourism context is a unique topic. Privacy is assumed to have low perceived importance, and tourists are less aware of it (Tussyadiah et al., 2019). Despite low awareness, concerns are bound to increase due to increased technological adoption (Femenia-Serra et al., 2022).

PCs have been defined as people’s perceptions of the risks and unfavourable effects of gathering and disseminating personal details (Malhotra et al., 2004). It also refers to how users feel about the accessibility and publication of their details (Ooi et al., 2018). The
current study encompasses all the PCs about tourism, including its impact on tourists and the strategies marketers adopt to curb them. Previous studies show that as technology progresses, privacy issues grow (Smith et al., 2011). Despite rising concerns, no comprehensive systematic literature review has synthesized PCs in tourism to the best of the author’s knowledge.

Buhalis and Amaranggana (2015) suggest that smart tourist destinations use big data to meet consumer preferences by offering the right services at the right time. It is believed that individualized services may increase users’ PCs (Habegger et al., 2014). Femenia-Serra et al. (2021) provide that smart tourism PCs stem from technology’s dangers, data misuse and data management incompetence. The current study conceptualizes PCs as people’s perceptions of the risks and unfavourable effects of gathering and disseminating personal details (Malhotra et al., 2004) and synthesizes the literature using privacy, security and data protection as the primary keywords. Most studies on privacy in tourism focused on technical development (Afolabi et al., 2020) or smart tourism (Gong and Schroeder, 2022). This study uses a hybrid topic modelling and bibliometric analysis approach to analyse scholarly literature on tourist industry PCs. Figure 2 shows the study’s research questions, methods and results.

This paper is structured as follows: Section 2 establishes the research gap, followed by Section 3 on methodology. Section 4 describes the analysis performed. Section 5 presents the content analysis, and Section 6 provides implications and future research directions. Finally, Section 7 concludes the research.

2. Establishing the research ground

Systematic literature reviews must incorporate transparency and rigour (Tranfield et al., 2003). Hall and Ram (2020) provided their perspective on privacy in tourism and distinguished privacy from security. The views on PCs in tourism were fragmented. The past reviews analysed a limited set of articles and have imposed geographical boundaries, leading to a myopic view of PCs. Figure 1 maps the identified research gap. The current study tries to address these identified research gaps by adopting a rigorous scientific machine-learning method to ensure more comprehensive and bias-free coverage of the literature.

![Figure 1](image-url)  
**Figure 1** Limitations in the existing studies on privacy concerns in tourism

*Source: Figure by authors*
The systematic review aims to achieve the following primary research objectives:

R01. To identify the privacy concerns (PCs) pertinent to the travel and tourism industry using robust scientific techniques.

R02. To identify the lacunas in the extant literature and propose future research agenda based on these gaps.

Researchers have applied a similar Methodology (Duong et al., 2022). Our primary contribution is to draw a broad picture of PCs in tourism that integrates recent findings with potential future directions. Figure 2 provides a general overview and maps the research questions, methodology and outcomes.

3. Research methodology

The proposed methodology framework has been adopted from Duong et al. (2022) to conduct the systematic literature review.

3.1 Selection of articles

Most core academic material is published in peer-reviewed journals, wherein Web of Science and Scopus indexed are the most pertinent ones (Moro et al., 2019). Since Scopus has a more extensive database and covers more titles than other databases (Srivastava et al., 2022; Thakral et al., 2023), it was selected to search for publications on privacy issues in the tourism industry. The aim is to have comprehensive coverage, i.e. from the publication of the first article to 2022, to cover a sizeable portion of recently published research on PCs.

The initial search query resulted in a sample of 11,175 papers (Figure 3). Such a wide range of publications reflects the diversity of the themes surrounding PCs and a range of industries, including health, food, sports, border security, smart tourism, aviation and transportation. We looked for journal articles only on privacy issues written in the English language and published in the categories of business, management or computer science only as listed in Scopus to compile the complete sample. We determined the search terms using the method Mustak et al. (2021) described and looked up the key terms used in other reviews (e.g. Mariani and Baggio, 2021). The search string used for data collection was as follows:

![Figure 2](image_url)

**Figure 2** Research outline of the study

- **Research Objectives**
  - R01. To identify the significant privacy concerns (PCs) pertinent to the travel & tourism industry as a whole using robust scientific techniques. (Gap 1 & 2)
  - R02. To identify the existing gaps in the body of knowledge and identify future research agendas based on these gaps.

- **Research Design and Methodology**
  - STEP 2: Fetched data of 2005 documents
  - STEP 3: Cleaned data to obtain 1073 articles suitable for further analysis
  - STEP 4: Clustered documents into four clusters using the topic modeling approach.

- **Research Outcomes**
  - Privacy literature in tourism can be broadly classified into four categories
  - 1) Privacy preservation & algorithm development
  - 2) Adoption of Public services & cybersecurity concerns
  - 3) Sustainable tourism and Governance
  - 4) Consumer behavior & tourism marketing
  - PCs in tourism can be attributed to various technologies including IoT, AI, block chains and other smart travel applications
  - Due diligence should be taken to preserve privacy of travelers as it can impact growth, economic benefits and sustainability of tourism.

**Source:** Figure by authors
Figure 3  Research architecture adopted in the study

Source: Figure by authors

(TITLE-ABS-KEY ("Privacy" OR "Security" OR "Data-protection") AND ("Travel" OR "Tourism" OR "Hotel" OR "Hospitality") AND (LIMIT-TO (DOCTYPE, "ar"))) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "BUSI")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j"))
After applying all the inclusion and exclusion criteria, 2,095 articles were received. This data set is cleaned manually by reading the title, abstracts and keywords. Articles specifically talking about PCs were included in the final data set, whereas articles focusing on job security, food security, airport and border security, terrorism and security concerning COVID-19 were not included in the final data set. This extensive data cleaning resulted in 1,073 articles for the final analysis.

3.2 Descriptive analysis

Figure 4 displays the distribution of the papers about PCs in tourism. Statistics reveal that academicians have paid significant attention to PCs in the past 10 years, which have accelerated exponentially. Around 634 articles were published in the previous five years (2018–2022), accounting for 59.32% of the total works published during the study period (1979–2022). One hundred nine publications have already been published within the two quarters of 2022, depicting rising scholarly interests.

Table 1 lists the top 20 publications in the data set contributing to the privacy research domain in tourism, totalling 218 papers or 21% of all publications produced. Most published works were in the high-ranked ABDC or ABS list. The publication of articles in various journals from different fields demonstrates the broad scope of privacy research. These journals might be a potential outlet for future studies in this domain.

Table 2 provides the top ten countries that have contributed the maximum to the domain of privacy research in tourism. The country’s contribution has been assessed based on its published documents. The country-wise contribution provides that the domain of PCs in tourism is mainly explored in developed nations. Thus, it allows scholars from developing or underdeveloped nations to explore the domain further. New geographies can offer newer insights into their PCs. The USA (260 documents, 9,273 citations) has contributed the maximum to the domain, followed by China (150 documents, 2,893 citations) and India (132 documents, 1,043 citations).

Table 3 lists the top ten most cited articles based on their citations. These articles guide future scholars about the field and topics with maximum acceptance and usability. Also, the statistics established the immense potential privacy research has in the tourism industry. However, the citations mentioned did not account for the age of the article, and hence, likely to give more weightage to articles published earlier.

Figure 4 Distribution of publications across the study period (1979–2022)

Source: Figure by authors
Table 1  Top 20 contributed journals across the data set period (1979-2022)

<table>
<thead>
<tr>
<th>Journals</th>
<th>No. of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Journal of Contemporary Hospitality Management</td>
<td>27</td>
</tr>
<tr>
<td>International Journal of Hospitality Management</td>
<td>25</td>
</tr>
<tr>
<td>Tourism Management</td>
<td>21</td>
</tr>
<tr>
<td>Journal of Hospitality and Tourism Technology</td>
<td>17</td>
</tr>
<tr>
<td>Journal of Travel Research</td>
<td>14</td>
</tr>
<tr>
<td>Tourism Management Perspectives</td>
<td>13</td>
</tr>
<tr>
<td>IEEE Transactions on Intelligent Transportation Systems</td>
<td>12</td>
</tr>
<tr>
<td>Cornell Hospitality Quarterly</td>
<td>9</td>
</tr>
<tr>
<td>Journal of Hospitality Marketing and Management</td>
<td>9</td>
</tr>
<tr>
<td>Journal of Hospitality and Tourism Management</td>
<td>8</td>
</tr>
<tr>
<td>Tourism Review</td>
<td>8</td>
</tr>
<tr>
<td>Journal of Hospitality and Tourism Research</td>
<td>7</td>
</tr>
<tr>
<td>Journal of Vacation Marketing</td>
<td>7</td>
</tr>
<tr>
<td>Current Issues in Tourism</td>
<td>7</td>
</tr>
<tr>
<td>Computers and Security</td>
<td>7</td>
</tr>
<tr>
<td>Service Industries Journal</td>
<td>6</td>
</tr>
<tr>
<td>Journal of Advanced Transportation</td>
<td>6</td>
</tr>
</tbody>
</table>

Source: Table by authors

Table 2  Country-wise contribution

<table>
<thead>
<tr>
<th>Country</th>
<th>Documents</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>260</td>
<td>9,273</td>
</tr>
<tr>
<td>China</td>
<td>150</td>
<td>2,893</td>
</tr>
<tr>
<td>India</td>
<td>132</td>
<td>1,043</td>
</tr>
<tr>
<td>UK</td>
<td>66</td>
<td>1,468</td>
</tr>
<tr>
<td>Australia</td>
<td>61</td>
<td>1,730</td>
</tr>
<tr>
<td>Malaysia</td>
<td>49</td>
<td>529</td>
</tr>
<tr>
<td>South Korea</td>
<td>47</td>
<td>1,439</td>
</tr>
<tr>
<td>Spain</td>
<td>45</td>
<td>1,192</td>
</tr>
<tr>
<td>Taiwan</td>
<td>42</td>
<td>1,070</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>36</td>
<td>1,596</td>
</tr>
</tbody>
</table>

Source: Table by authors

Table 3  Most cited articles

<table>
<thead>
<tr>
<th>Author</th>
<th>Journal title</th>
<th>Cited by</th>
<th>h-core author</th>
<th>h-core journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curbera et al. (2002)</td>
<td>IEEE Internet Computing</td>
<td>956</td>
<td>–</td>
<td>116</td>
</tr>
<tr>
<td>Bart et al. (2005)</td>
<td>Journal of Marketing</td>
<td>877</td>
<td>15</td>
<td>253</td>
</tr>
<tr>
<td>Kim et al. (2011)</td>
<td>Tourism Management</td>
<td>393</td>
<td>33</td>
<td>216</td>
</tr>
<tr>
<td>Ponte et al. (2015)</td>
<td>Tourism Management</td>
<td>329</td>
<td>18</td>
<td>216</td>
</tr>
<tr>
<td>Zhang et al. (2019)</td>
<td>Transportation Research Part C: Emerging Technologies</td>
<td>327</td>
<td>7</td>
<td>147</td>
</tr>
<tr>
<td>Hoh et al. (2006)</td>
<td>IEEE Pervasive Computing</td>
<td>270</td>
<td>17</td>
<td>103</td>
</tr>
</tbody>
</table>

Source: Table by authors
3.3 Pre-processing

Pre-processing removes unnecessary information. This stage has several processes, but tokenization, stop word removal and lemmatization techniques are essential for preparing the article abstracts. A similar methodology was used by Duong et al. (2022). Data were first converted into lowercase. Tokenization is done to break down large chunks of text into smaller tokens. Also, some words do not significantly contribute to the topic under examination. Hence, they should be removed. This must be done post-tokenization and helps reduce the data set’s size. Finally, lemmatization is done; the inflected words are condensed in this phase. It helps in retrieving the appropriate and required terms.

3.4 Latent Dirichlet allocation-based clustering and results

A standard method for classifying articles into subjects (also known as clusters) based on their semantic similarity is Latent Dirichlet allocation (LDA-based topic modelling) (Blei et al., 2003). Creating a latent layer, which reduces the number of dimensions between words and associated documents, is the objective of LDA. In this generative probabilistic model, each subject is assumed to be a combination of a base set of words, and each document is assumed to be a combination of a base set of topic probabilities. The clusters to which publications are assigned will also be the SE latent groupings. LDA likewise provides the most popular and representative terms for each cluster. Before clustering, LDA mandates that the user gives the number of groups (K). Blei et al. (2003) first suggested a perplexity score to assess how well an LDA model could extract meaningful clusters. Further, Syakur et al. (2018) used a more refined approach using the elbow method wherein the sum of squared errors (SSE) was used to calculate the value of K [1]. The point where SSE changed substantially, causing a change of angle. As evident in Figure 5, the SSE changed when K = 4; hence, we selected four as the ideal number of clusters.

The processes followed in clustering and labelling our data set are the following:

- Abstract preparation and cleaning.
- Finding the ideal K value using the elbow method approach.
- Using the ideal K and the LDA model to cluster 1,073 abstracts.
- Using the LDAvis approach, visualizing the top terms in each cluster (Sievert and Shirley, 2015).
- Labelling of clusters.

Figure 5 Sum of squared error (SSE) and the number of clusters value

Source: Figure by authors
The top 20 keywords for each cluster are included in Table 4 and the articles for each cluster. We present the semantic distance map between the keywords in the four clusters and the top keywords in each cluster using LDAvis (Sievert and Shirley, 2015). The trained LDA model is sent into LDAvis to visualize keywords. The outcomes of LDAvis are shown in Figure 6 for one cluster. Results from LDAvis can be divided into two categories:

1. The semantic map is demonstrated over the clusters on the left side. In terms of semantic words, four clusters stand out the most.

2. The top terms in each cluster are displayed on the right side, with the value adjusted by $\lambda$.

The value of $\lambda$ ranges from 0 to 1. When the value of $\lambda$ is close to 1, it provides the set of keywords shared among all the clusters. However, when the $\lambda$ values touch 0, it provides only keywords unique to a particular cluster. The 20 leading terms are shown in Table 4 in four groups with $\lambda = 1$ and $\lambda = 0$.

We used the most frequently mentioned terms for each cluster in Table 4 and the knowledge of our research team to label each cluster. The clusters are individually labelled by each researcher in the group, followed by a group discussion and cross-checking to arrive at the final labels for the four clusters. As a consequence, Table 5 shows the labels for the four clusters.

### 4. Bibliometric analysis for content analysis articles

The next step is to select the most representative articles for each of the four LDA clusters. We used bibliometric research as it is scientifically sophisticated, quantitative and extensively used (Si et al., 2019). It is commonly used to choose articles for content analysis in privacy research and tourism (Johnson and Samakovlis, 2019). We used VOSviewer to conduct the analysis (van Eck and Waltman, 2014). The study team used its vast

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Top 20 keywords with $\lambda = 0$ and $\lambda = 1$ values for individual clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0$</td>
<td>$\lambda = 1$</td>
</tr>
<tr>
<td>networks</td>
<td>privacy</td>
</tr>
<tr>
<td>IoT</td>
<td>proposed</td>
</tr>
<tr>
<td>algorithm</td>
<td>location</td>
</tr>
<tr>
<td>protocols</td>
<td>information</td>
</tr>
<tr>
<td>scheme</td>
<td>users</td>
</tr>
<tr>
<td>privacy-preserving</td>
<td>security</td>
</tr>
<tr>
<td>encryption</td>
<td>travel</td>
</tr>
<tr>
<td>nodes</td>
<td>results</td>
</tr>
<tr>
<td>sensor</td>
<td>devices</td>
</tr>
<tr>
<td>queries</td>
<td>attack</td>
</tr>
<tr>
<td>trajectory</td>
<td>road</td>
</tr>
<tr>
<td>vehicles</td>
<td>mobile</td>
</tr>
<tr>
<td>simulation</td>
<td>performance</td>
</tr>
<tr>
<td>packet</td>
<td>services</td>
</tr>
<tr>
<td>tracking</td>
<td>system</td>
</tr>
<tr>
<td>message</td>
<td>network</td>
</tr>
<tr>
<td>vanet</td>
<td>secure</td>
</tr>
<tr>
<td>monitoring</td>
<td>time</td>
</tr>
<tr>
<td>data sets</td>
<td>scheme</td>
</tr>
<tr>
<td>experiments</td>
<td>communication</td>
</tr>
</tbody>
</table>

Source: Table by authors
Figure 6  Semantic keyword visualization with \( \lambda = 1 \) (A) and \( \lambda = 0 \) (B)

Source: Figure by authors

Table 5  Cluster labelling and publications results of LDA

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Label</th>
<th>No. of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Privacy preservation and algorithm development</td>
<td>324</td>
</tr>
<tr>
<td>2</td>
<td>Adoption of Public services and cybersecurity concerns</td>
<td>300</td>
</tr>
<tr>
<td>3</td>
<td>Sustainable tourism and governance</td>
<td>172</td>
</tr>
<tr>
<td>4</td>
<td>Consumer behaviour and tourism marketing</td>
<td>278</td>
</tr>
</tbody>
</table>

Source: Table by authors
experience and professional opinion to label topic modelling results (Table 5). However, we did a reliability analysis to guarantee that the study team had a consensus and that the content analysis was neutral. Inter-rater reliability was used for analysis (O'Connor and Joffe, 2020). Data are independently coded and compared for agreement (Krippendorff, 2004). A small sample of ten articles was used, which was more than 10% of the selected data set and is considered sufficient for the analysis (Krippendorff, 2004). We assessed inter-coder reliability using Miles and Huberman (1984). Inter-coder reliability was 0.82, indicating good agreement among coders (Landis and Koch, 1977).

4.1 Co-citation and bibliographic coupling analysis

Co-citation (Small, 1973) and bibliographic coupling (Kessler, 1963) analysis are two scientific methods to map the citation links between publications systematically. If the papers cite or are mentioned by each other more frequently, the likelihood that they reflect the same clusters and approaches increases (Hjørland, 2013). Co-citation and bibliographic coupling connect two publications with a third document; as a result, the timing has no bearing on the relationship because the third paper serves as a complement. We condense the definitions of the two analyses: When paper A references both papers B and C, it is said that these papers are co-cited; however, when both papers B and C quote paper A, they are displayed to be bibliographically linked (Van Eck and Waltman, 2014). The well-discussed publications inside a cluster may be found using these methods, which have been widely used in various fields (Modak et al., 2020).

In this study, we use a hybrid technique, simultaneously executing co-citation and bibliographic coupling analyses for each cluster. We establish a threshold of three connections to be included in the network and input four clusters to VOS viewer. The papers that meet both criteria are selected:

1. at least three linkages with co-citations; and
2. have at least three connections between bibliographies.

Consequently, 68 publications from Cluster 1, 63 papers from Cluster 2, 36 articles from Cluster 3 and 58 papers from Cluster 4 are chosen for additional analysis. There is a concern that the recent articles might get missed in the above analysis. Therefore, we examined articles published between 2019 and 2022 to identify current trends and added 132 articles across the four clusters. Thus, we had Cluster 1 (98 papers), Cluster 2 (91 papers), Cluster 3 (58 papers) and Cluster 4 (110 papers), equating to a total of 357 papers. Finally, to ensure the data set’s quality, we only analysed those articles that are at least ABDC/ABS listed. The findings indicate that 68 out of 357 publications were published in ABDC/ABS-ranked journals. Table 6 provides characteristics for each cluster.

4.2 Analyzing cluster using co-occurrence keyword heatmap

In this section, publications undergo a co-occurrence keyword analysis. The analysis summarizes article topics and determines sub-clusters for each cluster. The research uses VOSviewer. Density presentation in heatmap form lets the programme preview article content by cluster. Cluster 1 has 611 terms; Cluster 2 has 515; Cluster 3 has 266; and Cluster 4 has 336. A glance at Cluster 1 shows how to avoid privacy breaches and protect personal data [Figure 7(a)]. “Data privacy”, “optimization”, “algorithms”, “security systems”, “dataset” and “privacy-preservation” are the most frequently used terminology. Cluster 1 articles emphasize privacy and optimizing privacy models and algorithms. Figure 7(a)'s borders show privacy-preserving alternatives for automobiles and transportation.

“Technology adoption” stands out in Cluster 2 [Figure 7(b)]. “Ride-sharing”, “air travel”, “transport” and “crime” are common phrases. This indicates that the authors focus on
Table 6  No. of selected publications for content

<table>
<thead>
<tr>
<th>Clusters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected papers with all connections</td>
<td>15</td>
<td>11</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Recently published papers without connections</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Source: Table by authors

Figure 7  Density graphs for Cluster 1 (a), Cluster 2 (b), Cluster 3 (c) and Cluster 4(d)

Source: Figure by authors
public services and cyber security issues related to their adoption. “Hygiene consciousness”, “satisfaction” and “service quality” are service evaluation phrases. It means services are examined on macro and micro levels, and cyber security assaults’ potential losses have been discussed. Cluster 3 discusses tourist sustainability, corporate social responsibility and system governance. The cluster has few keywords, as seen in Figure 7(c), indicating less privacy study on tourism. Cluster 4 includes extensive tourist privacy studies. Tourism practises and individual behaviours are included. “Trust”, “customer satisfaction”, “loyalty”, “guest experience”, “hotel attributes” and “tourism market” are the most important keywords.

5. Content analysis

All the selected articles were examined in full, and the insights of each of the four clusters’ sub-themes are thoroughly explored in this part. Our outcomes reveal some major research issues that merit further examination compared to earlier literature reviews on similar topics, such as Tussyadiah et al. (2019). Below are the specifics of our research’s central themes.

5.1 Cluster 1

Cluster 1 ranked as the second highest cluster, with 19 articles. The articles in this particular segment majorly talk about the security of data and preserving privacy using the latest technologies. PCs have received extraordinary attention in tourism, especially smart tourism. As a result, scholars have also looked at the solutions for the concerns arising, which resulted in the development of specific applications and algorithms to protect users’ privacy. The papers within this cluster can be segregated into the following sub-themes:

5.1.1 The personalization privacy-paradox. Marketers customize their services using client data. Individualized services improve user experience and tourist e-satisfaction (Mahadin et al., 2020). To do this, companies need tourist data. It highlights a crucial paradox for customization businesses: tourists prioritizing information integrity are less likely to customize (Awad and Krishnan, 2006). However, people today want convenience (Klaus and Zaichkowsky, 2021), which requires customized solutions demanding frequent information disclosures. More privacy-conscious tourists may think tailored services are less beneficial to clients (Stone et al., 1983). It is observed that hotel visitors divulge more information by using apps for personalized services (Morosan and Defranco, 2016). Online travel businesses need quality content to satisfy customers (Kim and Lee, 2005). Thus, travel agencies acquire more tourist data to improve their services. It is found that assuring privacy increases service usability, reducing consumer privacy worries (Lee and Cranage, 2011).

5.1.2 Algorithm and model development. Since PCs significantly impact purchase intent directly and indirectly through trust, the tourist sector developed several strategies to address them (Eastlick et al., 2006). Travellers can also use beautiful, convincing smartphone apps to make decisions. Initial studies focused on internet of things (IoT) as a significant privacy risk and developed solutions like vehicular ad hoc network (VANET) (Schoch et al., 2008) for smart interconnected automobiles. Tourism innovations, including end-to-end-encrypted mobile platforms for real-time information connected stakeholders (Gretzel et al., 2020), and privacy-preserving technologies failed to meet privacy issues. Tourists are more concerned about privacy when booking tickets, travel applications and location-based services (Watkins et al., 2018). Internal conflicts affect the privacy-preserving mechanism, while technology insecurity and consumer trust affect user use intentions (Lee and Rha, 2016).
5.2 Cluster 2

With the rising demand for tourists, Cluster 2 emerged as a significant research area on the association between PCs and tourism. With 13 ABDC-listed journal articles, Cluster 2 is in the bottom two. The initial analysis provides that the articles in this cluster talk about PCs emerging from the user adoption of smart public services and cyber security attacks. This cluster is further segregated into two sub-themes:

5.2.1 Tourists’ adoption of smart public services. Tourists’ need for more extraordinary public amenities has raised privacy problems. Privacy and security concerns affect e-government service adoption (Wong et al., 2011). Tourists’ adoption of smart public services such as air travel services (Morosan, 2014), e-passports (Meingast et al., 2007) and intelligent transportation (Lederman et al., 2016) have raised PCs. Such services need users to disclose sensitive personal information, causing PCs. However, IoTs have caused a digital revolution in tourism, capturing numerous data points daily but generating tourist privacy worries (Omolara et al., 2022).

5.2.2 Cyber-security issues. Several high-profile organizations in the sector have garnered undesired attention due to their fragmented, technology-focused and compliance-oriented cybersecurity approach (Paraskevas, 2022). Kaspersky Lab (2018) estimates security breach recovery costs over $500,000. Tourism and hospitality companies must improve privacy beyond compliance-based information governance. Ethical data solutions also get a competitive edge (Yallop and Seraphin, 2020). The tourism sector is confused and duplicated by a complicated network of public tourist authorities wherein the rules are still ambiguous (Devine and Devine, 2011).

5.3 Cluster 3

Cluster 3 was the most emerging field in the privacy domain. However, it is the smallest cluster with 11 articles. The results of keyword analysis and LDA provides that the articles under this segment talk of sustainability in tourism and the changes in the governance of tourism and privacy policies. A bird’s eye view of the results provides that this cluster focuses on sustainable tourism concepts such as corporate social responsibility and other avenues and discusses their impact on users’ PCs.

5.3.1 Sustainability and privacy concerns. Sustainable tourism has been popular among tourists (Demunter, 2017). Tourism sustainability affects privacy in many ways. Tourism sustainability may be enhanced by social media use (Martínez-Navalón et al., 2020). It may raise PCs. Sustainable or eco-tourism explores green places, which raises location-based privacy problems (Prashyanusorn et al., 2010). Sustainable tourism research aims to mitigate tourism’s negative impact on the environment, society and economy. Using big data has offered several impressive solutions. These tactics do not appear to be applied to sustainable tourism (Rahmadian et al., 2022). Privacy-themed social dinners are believed to help sustain tourism (Davies et al., 2022). Data cost, quality and privacy remain issues for big data research (Li et al., 2018). Big data may cover many high-quality possibilities, but not all (Xu et al., 2020). User-generated content and internet reviews are also unreliable, with false reviews hurting users (Xu et al., 2020). Despite its usefulness, cell phone roaming and transaction data have not been used in tourist research due to privacy issues.

5.3.2 Governance-related issues. Tourism governance is hampered by unclear privacy and data usage rules. Tourism and hospitality organizations (THOs) use big data and analytics to inform business decisions. Data analysis helps THOs survive and revive the tourism business during global crises and ambiguities. Big data and digital technology bring economic growth, but ethical, privacy and security challenges arise (Yallop et al., 2021). Tourism governance should consider other important aspects of privacy and ethics, a fair exchange of traveller data, and THOs’ ability to show a social licence to operate by building trust with stakeholders. Governance-related issues are more prominent in smart cities and...
public services where users’ data is traded frequently (Lee et al., 2020). The tourism industry is increasingly looking towards blockchain technology to resolve governance-related issues (Tyan et al., 2020).

5.4 Cluster 4

The last segregation made, i.e. Cluster 4, consists of 25 articles and is the largest cluster. The cluster includes a set of homogeneous research articles examining the individual-centred phenomena in tourism, such as the tourist’s perception of hotel service quality and PCs. Similarly, traveller’s satisfaction, intention and loyalty amidst rising PCs were also examined. The articles also focus on business strategies and tourism marketing practices leading to PCs. Cluster 4 is segmented into the following sub-themes.

5.4.1 Consumer behaviour and individual-centred privacy concerns. Privacy literature has focused on individuals’ reactions to tourism technology. Several individual-centred phenomena have been examined so far, such as travellers’ trust in biometric technologies (Pai et al., 2018), satisfaction and loyalty to website design and travel apps (Morosan and Defranco, 2016). PCs and self-efficacy influence tourists’ use of specialized technology, such as smartphone apps (Wozniak et al., 2018). Kim and Qu (2014) explored tourists’ PCs and service innovation intentions. Past studies used the ease of use and usefulness to differentiate business and leisure tourists (Zhang et al., 2019). Finally, the literature looked at enhancing the tourism experience. It explores the factors that influence tourists’ use of technology. Hardware and software, demands and contexts, usability and supply and connectivity are found to be the major ones (Tanti and Buhais, 2017). Another study discovered that facilitating conditions, habit and buy intention is essential to passengers’ actual behaviour while purchasing online trips (Dogra and Adil, 2022). However, there are still recent advancements in tourism that require further examination.

5.4.2 Tourism marketing and business strategies led to privacy concerns. Tourism marketing and commercial strategies raise privacy problems. Websites often advertise third-party privacy certifications to gain customers’ trust (Chen et al., 2021). When corporations approach users’ personal information as a commodity (Malhotra et al., 2004), privacy breaches and data usage concerns occur. Companies need tourists’ data to sell tourism and customize services (Lee and Cranage, 2011), raising PCs. Privacy considerations affect opinions towards social media ads (Rana and Arora, 2022). Hotels’ methods for mitigating consumer moving infractions and privacy issues have been investigated (Yu et al., 2022). Also, it is found that PCs can play a significant role in promoting medical tourism (Zolfagharian et al., 2018).

5.5 The conceptual framework

The conceptual framework in Figure 8 summarizes the outcomes of the cluster analysis, co-citation analysis and content analysis. The framework in Figure 8 explains the catalyst for the rise of privacy issues in the tourism sector. The tourists increasing desire to have better public services across destinations, the willingness to develop sustainable tourism through advanced technologies, and the desire to reach and target the vast consumer markets and the efforts to enhance tourist satisfaction and experience led the tourism industry to increasingly adopt advanced technologies in haste leading to rise in PCs of the individuals. Further, they are found to have a two-way relationship. The desire of consumers to have more convenience (Klaus and Zajichowsky, 2021) forced tourism to adopt technology, which, in turn, led to PCs. Simultaneously, the increased PCs forced the tourism industry to develop more robust technologies.

6. Implications and future research directions

The current analysis helps tourism stakeholders make educated privacy decisions. Firstly, tourists drive the industry; this literature synthesis helps them make appropriate disclosure
decisions. Tourists should be cautious when using public services and avoid overly convenient and customized services. Secondly, administrators of intelligent tourism locations should preserve visitors’ personal information since privacy issues affect tourists’ visit intentions. Managers must realize that technology helps achieve sustainability but increases privacy infringement. Thirdly, the study shows how authorities and regulating bodies protect privacy (Zolfagharian et al., 2018).

The study mandates specific policy changes, such as the inclusion of privacy preservation in the definition of sustainable tourism and the development of a rating scale for tourism destinations based on their privacy-preserving measures. The study provides tourism marketers with a new promotional tool by advertising their privacy preservation ratings. It establishes the impact of privacy preservation in enhancing the tourist’s willingness to visit a destination. Finally, the article provides directions for future researchers. Future scholars can synthesize the literature with a framework-based analysis such as theory-context-characteristics-methodology (Paul et al., 2021) or other theoretical frameworks. Bibliographic coupling and co-citation analysis were considered suitable for the bibliometric research. However, future studies can use a more robust methodology. Regarding the selection criteria, only ABDC/ABS-ranked journals were considered for the final content analysis. Future scholars can use other quality analysis criteria to get to the final data set. Further, the cluster-wise future research directions are provided in Table 7.

7. Conclusion

This study reviews travel-related privacy research and suggests future studies. The study involves 1,744 publications from Scopus. The review framework performs machine learning-based clustering and bibliometric analysis. The content analysis included 68 publications. Four clusters were notable. Cluster 1 explained “Privacy preservation and algorithm development”, Cluster 2 “Smart public services and cyber-security”, Cluster 3 “Sustainable tourism and governance” and Cluster 4 “Consumer behaviour and tourist marketing”. The study summarizes privacy-related research themes. Analyzing contemporary tourism privacy literature helps researchers locate their studies on our privacy literature map. It facilitates future research in their field. The outlined four study directions serve as a foundation for potential studies on privacy issues in the tourism industry. This study has
several limitations. Choosing keywords for search queries cannot be exhaustive to start with. Despite our best attempts to eliminate irrelevant papers, we think our data set may still contain some unavoidable publications when working with a large data set.

Note

1. The point from where the additional changes in SSE become minimal, i.e. where the curve shows the maximum bend, is chosen to be the ideal $K$.

References


Corresponding author
Sajjad M. Jasimuddin can be contacted at: sajjad.jasimuddin@kedgebs.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com