ARTICLE
Doi: 10.4025/rimar.v13i2.66702
Received: 01/18/2023
Revised: 04/14/2023
Accepted: 08/08/2023
Double blind review, SEER/OJS
Editor for this article: Francisco Vieira

DATA PRIVACY IN RETAIL: A BIBLIOMETRIC ANALYSIS

Privacidade de dados no varejo: uma análise bibliométrica

Alina Flores [1]
0000-0001-6162-4659
✉ alina.flores@ufrgs.br

Lucas Dorneles [1]
0000-0003-1904-9552
✉ lucas.dorneles@ufrgs.br

[1] *Federal University of Rio Grande do Sul*

ABSTRACT

Data privacy is an issue that is catching the attention of the retail research area, especially with the technological advances in data collection and analysis that impact the consumer's perception of safety. As publications on this subject have increased significantly in recent years, a review of academic knowledge is needed. In order to address this, our study conducts a comprehensive review of the privacy literature. Our investigation of the intersection between data privacy and retail literature is a bibliometric analysis. In this paper, 31 years of publications were collected from two sources of bibliographic databases: Web of Science and Scopus, with 413 articles in total. To facilitate our analyses, we use the Bibliometrix package in R language. Our analysis revealed that this is a subject growing in size and relevance, but the intellectual structure of the field is fragmented. We also discussed, based on the conceptual structure of the publications, ideas for future research regarding privacy and retail.

KEYWORDS: retailing; data privacy; bibliometric analysis; Bibliometrix; literature review.

RESUMO

Privacidade de dados é uma questão que está ganhando a atenção da área de pesquisa em varejo, especialmente com os avanços tecnológicos na análise e coleta de dados que impactam a percepção de segurança dos consumidores. Como as publicações deste assunto aumentaram significativamente nos últimos anos, uma revisão do conhecimento acadêmico se torna necessária. Com o objetivo de resolver essa questão, nosso estudo realiza uma revisão abrangente da literatura sobre privacidade de dados. Nossa investigação sobre a interseccionalidade entre as literaturas de privacidade de dados e varejo é feita por meio de uma análise bibliométrica. No presente trabalho, 31 anos de publicações foram coletados de duas bases de dados científicas: Scopus e Web of Science, com total de 413 artigos. Para facilitar a nossa análise, nós utilizamos o pacote Bibliometrix para linguagem R. Nossa análise mostrou que este é um assunto que está crescendo em tamanho e relevância, mas a estrutura intelectual do campo ainda é fragmentada. Nós também discutimos, baseados na estrutura conceitual das publicações, ideias de pesquisas futuras e apresentamos uma agenda de pesquisa para privacidade e varejo.

PALAVRAS-CHAVE: varejo, privacidade de dados, análise bibliométrica, Bibliometrix, revisão de literatura.

1 Introduction

Technological advancement is constantly bringing a wide range of business applications, such as the internet of things (IoT), big data, virtual reality, augmented reality, artificial intelligence, robots, drones, driverless vehicles, among others. Retail does not escape this reality, especially when it comes to options to better engage with consumers (Grewal, Roggeveen, & Nordfält, 2017). With more information, retailers are able to customize offers that, in concept, are very beneficial to both sides. However, one concern is growing: data privacy. An example of a breach of data privacy is the Cambridge Analytica scandal. In 2014, Facebook users' data was improperly collected without their consent and this data was allegedly used to influence the 2016's election (Confessore, 2018).

The personalization–privacy paradox discussed by Grewal et al. (2017) highlights that consumers can perceive how much of their data and information retailers have access to and this leads to a concern about their privacy. One way to address this is with the creation of some regulatory measures, such as the law nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais (LGPD), which is Brazilian legislation that regulates the activities of processing personal data, or EU Regulation nº 2016/679, General Data Protection Regulation (GDPR), which is the European Union to protect user data. The retail literature also addresses privacy concerns within different themes, such as artificial intelligence (Guha, Grewal, Kopalle, Haenlein, Schneider, Jung, Moustafa, Hegde, & Hawkins, 2021), big data (Dekimpe, 2020), retail strategy (Grewal, Gauri, Roggeveen, & Sethuraman, 2021), omnichannel (Cui, Ghose, Halaburda, Iyengar, Pauwels, Sriram, Tucker, & Venkataraman, 2021), consumer journey (Grewal & Roggeveen, 2020), online relationship (Eastlick, Lotz & Warrington, 2006), and even in an editorial in the Journal of Retailing (Roggeveen & Sethuraman, 2018).

Since 1975, we have had an annual percentage growth rate of 2.4% of articles published discussing privacy and retail together. But it's impressive to note that only the last 7 years represent more than 50% of the articles published in the area, which shows the growing relevance of the subject and the tendency to increase. This holds true in the business world. Forbes magazine released reports in 2019 and 2020 talking about the growing concern about the privacy of consumer data, including noting that this is becoming the most important issue of the next decade (Meehan, 2019; Goswami, 2020).

At this point, better understanding the state of knowledge about data privacy is very important for developing a theoretical perception of retailing. This study, therefore, conducts a comprehensive review of the privacy literature, with 2 specific objectives: (1) conduct a bibliometric review to analyze the research already carried out about privacy and retail; and (2) develop a research agenda identifying the aspects of privacy that could be subject of future research. The paper is structured as follows. First, we present the theoretical background, and briefly synthesize existing knowledge. Next, we describe the bibliometric analysis procedure adopted. We then discuss the results of the review findings for privacy. After that, we present our discussion and research agenda for future studies. Finally, we end with our conclusion.

2. Literature Review

Ethics is a complex construct involving several subconstructs (Guha et al., 2021). The focus of our study is on the dimension of customer security and, more specifically, on data privacy. It has a very clear tradeoff in the data privacy literature. With new technologies available to retailers, they are able to collect detailed data about each customer's journey across different touchpoints (Roggeveen & Sethuraman, 2018). This data easily turns into essential management information for a more personalized offer, even happening in real

time. However, customer privacy becomes a major concern. While there is a great effort to collect and analyze individual consumer-level data, security and privacy issues must also be addressed. One example is consumer concerns that marketers may misuse their private information. This is a clear case of the personalization–privacy paradox, and finding a balance is still a challenge (Grewal et al., 2017).

Phelps, Nowak and Ferrell (2000) created a conceptual model to understand consumer privacy concerns. As antecedent factors, the authors highlighted the type of personal information requested, the amount of control offered over their information, potential consequences and benefits, and consumer characteristics. This would lead to results such as a level of concern about how companies use personal data, and beliefs about marketing practices. Future outcomes such as purchase intent, use of direct marketing channels, support for rules that limit the collection and use of personal data, and request to remove your name from a mailing list were also mentioned and deserve attention. In this regard, consumer concerns about their privacy can affect their willingness to shop online, which has a big effect on e-commerce sales (Parasuraman & Zinkhan, 2002). Omnichannel strategies aim to provide customers with a cohesive and tailored experience across different contact channels. Nonetheless, it is imperative that consumers feel comfortable and secure engaging with these touchpoints at every stage of the purchase process (Verhoef, 2021).

According to Zhou, Lu and Ding (2020), the main challenge for retailers is “protecting customer privacy while gaining insights into customer data”. A frequent suggestion is that companies inform consumers, in advance, about their data usage policies and the benefits this generates for the shopper. This allows for a greater sense of control over the data itself and improves trust, which impacts purchase intentions (Culnan, 1995; Phelps et al., 2000). In support, Dekimpe (2020, p.7) says that “customer trust, which is essential to a retailer's success, takes time to build up, but can be lost very quickly if customers feel their data have been used without their permission, or have not been adequately protected”. It is important to emphasize that the feeling of invasion of privacy can vary according to different aspects, such as the recommendation system used by the company, product categories, among others.

The literature also mentions that a company's reputation leads to greater perceptions of trust among consumers, reducing the risks associated with privacy concerns (Eastlick et al., 2006). From a public policy perspective, the type of information collected and the amount of control consumers have over subsequent dissemination are the most relevant factors in thinking about privacy regulation (Eastlick et al., 2006).

Another key challenge for companies is determining how to adopt an omnichannel strategy while respecting consumer privacy. Cui et al. (2021) discuss that one possible way to address this issue is for the company to try to ensure that they do not predict behavior using only an individual's data or, if they do predict behavior at the individual level, try to ensure that such data are anonymous. It can be an initial path that allows companies to enjoy a large amount of data, without leaving the consumer too vulnerable or exposed. Still, on privacy, a curious aspect that caught our attention was Cui et al. (2021) suggestion that there could be a market for consumer data that would result in a fair assessment and at the same time preserve privacy, creating a win-win situation. Many consumers are increasingly willing to share their personal data with brands in exchange for some economic incentives (e.g. discounts). As the trend is toward more intelligent use of data to generate competitive information, it may be a cheap price to pay for valuable customer information. But future studies must consider the best win-win combination that the customer perceives as interesting, in order to align expectations.

The area of public policy plays a big role in helping to advance studies in the area of privacy. A clear suggestion is government intervention in the form of new regulations to protect consumer privacy, including the need for control agencies. A good example is the General Data Protection Regulation (GDPR) which was created with the purpose of providing a set of standardized data protection laws in all member countries of the European Union (EU). This should make it easier for EU citizens to understand how their data is being used and also to file complaints, even if they are not in the country where they are located. In Brazil, we have law nº. 13.709/2018 of GDPR in force. Bradlow, Gangwar, Kopalle and Voleti (2017) mention that retailers can address the ethical issue of customer privacy in some ways, for example, by enabling a clear acceptance policy for their customers regarding the collection and use of their data, showing the benefits of predictive analytics to their customer base, and rewarding customer loyalty.

3. Method

Concerning the main goal of this paper, we conducted a bibliometric literature review. This research method is widely used to analyze different results regarding a set of bibliographic documents. The bibliometric review analyses are very useful for providing an overview of academic research in an area, using statistical tools in order to identify the main trends in terms of publications, citations, authors, keywords, and institutions (Martínez-López, Merigó, Valenzuela-Fernández & Nicolás, 2018).

To ensure the representativeness and high quality of studies included in our review, we collected bibliographic documents from Web of Science and Scopus databases, which resulted in 260 and 342 articles, respectively. The search process used the terms “privacy”, and “retail” or “retailing” in the title, abstract, and keywords. We choose to filter articles only. For this review, we selected the search period from 1975 to 2022, but we found 31 years with articles published (from 1976 to 1983, from 1985 to 1986, from 1989 to 1992, from 1996 to 1998 there were no publications in our databases).

For the sake of more deeply analyzing the results, this paper used RStudio's Bibliometrix package for bibliometric investigation. RStudio has an open software nature that provides a wide variety of statistical and graphical techniques, which allows for good data visualization and analysis (Aria & Cuccurullo, 2017). Using Bibliometrix package we were able to run a descriptive analysis, with main information regarding the collection, for example, the most productive authors, the most cited papers, the most productive countries (based on first author’s affiliation), the most frequent journals, the annual scientific production, most relevant keywords, among others. In addition, we can visualize the intellectual structure of the area with co-citation analysis, and also the conceptual structure with cluster analysis.

Before starting the analyses, we pre-processed the articles database, removing the documents without authors and the duplicated ones. This resulted in a final database of 413 articles. The analysis showed that the 360 documents came from 270 different sources, which demonstrates a variety of journals addressing the topic.

Table 1 summarizes the data collection process along with the description of each one.

Table 1. Description of the data collection process

Data collection stages	Description
Date of search	December 1st, 2021
Databases	Scopus and Web of Science (WOS)
Search terms	“privacy”, and “retail” or “retailing”
Search focus	Title, abstract, and keywords
Refinement	Document type: Articles. Language: English.
Final database	413 articles

It becomes clear once we take a look at table 2 with the 10 most cited articles, that only the Journal of Business Research appears twice.

Table 2. The 10 most cited documents

R	TC	C/Y	Title	Author/s	Journal	Year
1	517	28.7	Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation.	Garbarino, E., & Strahilevitz, M.	Journal of Business Research	2004
2	388	24.2	Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment.	Eastlick, M. A., Lotz, S. L., & Warrington, P.	Journal of Business Research	2006
3	332	18.4	The theory of planned behavior and Internet purchasing.	George, J. F.	Internet Research	2004
4	328	16.4	Consumer perceptions of Internet retail service quality.	Janda, S., Trocchia, P. J., & Gwinner, K. P.	International Journal of Service Industry Management	2002
5	321	21.4	Role of electronic trust in online retailing.	Mukherjee, A., & Nath, P.	European Journal of Marketing	2007
6	264	20.3	Determinants of customer repurchase intention in online shopping.	Chiu, C., Chang, C., Cheng, H., & Fang, Y.	Online Information Review	2009
7	209	13.9	RFID: an enabler of supply chain operations.	Attaran, M.	Supply Chain Management: An International Journal	2007
8	202	40.4	Shopper-Facing Retail Technology: A Retailer Adoption Decision Framework Incorporating Shopper Attitudes and Privacy Concerns.	Inman, J. J., & Nikolova, H.	Journal of Retailing	2017
9	202	14.4	Users’ adoption of e-banking services: the Malaysian perspective.	Poon, W.	Journal of Business & Industrial Marketing	2007
10	202	20.2	A review of RFID technology and its managerial applications in different industries.	Zhu, X., Mukhopadhyay, S. K., & Kurata, H.	Journal of Engineering and Technology Management	2012

Notes: R = rank; TC = total citations; C/Y = citations per year
Source: Elaborated by the authors (2022)

But it is also interesting the results from table 3, which shows that four retail journals are among the 10 most relevant sources. Thus, we can see the importance of the topic of privacy for the area, which represented a total of 61 articles, with the Journal of Retailing and Consumer Services, and Journal of Retailing standing out.

Table 3. The 10 most relevant sources

R	Source	N
1	Journal of Retailing and Consumer Services	27
2	Journal of Retailing	14
3	International Journal of Retail and Distribution Management	12
4	IEEE Transactions on Smart Grid	8
5	International Journal of Electronic Marketing and Retailing	8
6	Internet Research	7
7	Journal of Business Research	7
8	International Journal of Online Marketing	6
9	Journal of Business Ethics	5
10	Applied Energy	4

Notes: R = rank; N = number os articles
Source: Elaborated by the authors (2022)

4. Data Analysis and Results

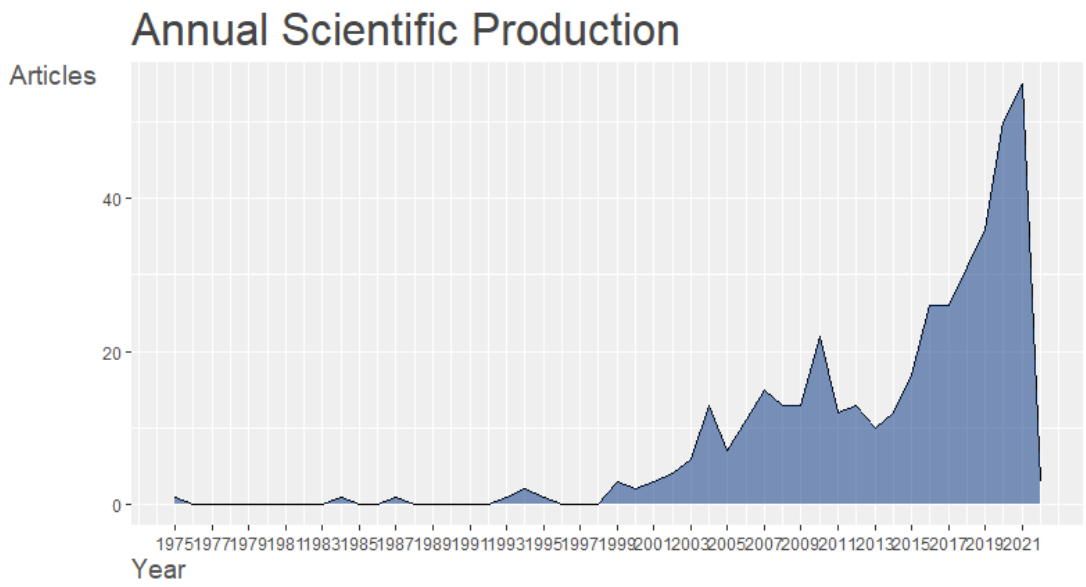
In this section, we will present the results from our bibliometrics, separating them into three subsections. First, we will bring the descriptives of the database, with the scientific production, most cited papers, and others. Then, we will discuss the intellectual structure of the field with a co-citation analysis. After that, we will examine the conceptual structure with a co-word analysis.

4.1. Descriptive analysis

We start our descriptive analysis by comparing our database through the lens of Lotka's law (Lotka, 1926). Lotka's law is used to compare the frequency distribution of a scientific field. It uses an inverse square law to describe the frequency of publication. In other words, it is known that authors produce a limited number of articles, and Lotka created a statistical law to calculate the frequency distribution of any scientific database. The theoretical beta coefficient of Lotka's law is 2. Any scientific publication database that is normally distributed should not have a beta coefficient statistically different from 2. Our database has a beta coefficient of 3.81 with a goodness fit equal to 0.98. The p-value of the Kolmogorov-Smirnoff two-sample test is 0.139, which indicates that there is no significant difference between the theoretical beta coefficient and ours. This indicates that our empirical database has similar productivity as the theoretical one.

Moving further, we can observe a growing tendency of scientific publications with concern for privacy and retail, just by looking at Figure 1. From 1975, the first year with records in our database, to 2010, 119 articles were published. In the last decade, from 2011, 291 articles were published, with 2021 being the most productive year, with 55 published articles so far. The annual percentage growth rate is 2.37, which also shows this increase in scientific production.

Figure 1. Annual Scientific Production from 1997 to 2022

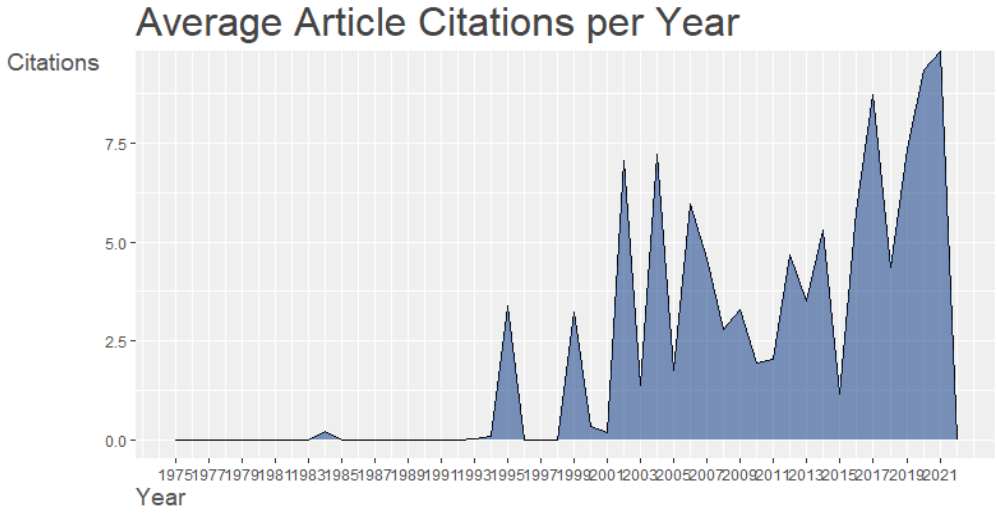


Source: Elaborated by the authors (2022)

The average article citation per year is also increasing (as portrayed in Figure 2). This, along with the increase in the number of published articles shows that the area is growing in size and relevance. The most productive country is the United States, with a total of 105 articles, 32.11% of the database, with 100 being published only by American authors (Single Country Publications) and 5 being published with authors from other countries (Multiple Country Publications). The next most productive country is the UK, but with less than 1/3 of the US production, with 27 published articles. The list of the 10 most productive countries is in Table 4.

For citations, the same pattern happens, with the most productive countries being the most cited ones. One highlight is Korea, which is not in the top 10 productive countries but is the fifth country with the most citations. Italy and Canada are very productive, but not very relevant, as they appear in the most productive countries, but not in the most cited ones.

Figure 2. Average Article Citations per year



Source: Elaborated by the authors (2022)

Table 4. Number of articles published by country for the first 10 countries

R	Country	Art.	Freq	SCP	MCP	MCP_Ratio	R	Country	TC	AAC
1	USA	105	0.32	100	5	0.0476	1	USA	4709	44.85
2	UNITED KINGDOM	27	0.08	25	2	0.0741	2	UNITED KINGDOM	599	22.19
3	CHINA	24	0.07	19	5	0.2083	3	CHINA	509	21.21
4	INDIA	22	0.07	22	0	0	4	GERMANY	494	26.00
5	GERMANY	19	0.06	17	2	0.1053	5	KOREA	327	54.50
6	AUSTRALIA	15	0.05	12	3	0.2000	6	MALAYSIA	321	40.12
7	ITALY	9	0.03	9	0	0	7	SPAIN	283	35.38
8	MALAYSIA	8	0.02	8	0	0	8	AUSTRALIA	274	18.27
9	SPAIN	8	0.02	8	0	0	9	FRANCE	246	61.50
10	CANADA	6	0.01	6	0	0	10	INDIA	241	10.95

Notes: R = Rank; Art. = Number of published articles; Freq. = % of the articles from our database; SCP = Single Country Publications; MCP = Multiple Country Publications; TC = Total Citations; AAC = Average Article Citation
Source: Elaborated by the authors (2022)

As mentioned, looking at table 2, we can see the top 10 articles with the most citations in our database. The most cited article was written by Garbarino & Strahilevitz (2004), has 517 citations, and was published in the Journal of Business Research (JBR). This article was from the beginning of the popularization of the internet and dealt with gender differences in trust from buying online. Although data privacy was not the central theme, it was relevant for the research and they showed that even when controlling for internet usage, women are more concerned with privacy loss on the internet when buying online (Garbarino & Strahilevitz, 2004). The second most cited article is also from JBR and it was published in 2006 by Eastlick, Lotz, and Warrington. This article was about applying a relationship marketing framework, that was used mostly in business-to-business, to business-to-consumer online retail. They found that trust and commitment, which were mediators in business-to-business relationship marketing, also act as mediators in business-to-consumer retail (Eastlick et al., 2006). Although it was published in 2006, it deals with very actual subjects, like privacy concerns in online retailing. As we can see in table 1, nine of the ten most cited articles deal with online retail or technology-related issues. This shows that privacy is more relevant when dealing in online retail contexts.

Another important aspect when looking at bibliometrics is who are the most productive authors. In our database, the most productive authors are Viswanath Venkatesh and Yi Wang, with 6 articles each. However, they are the first author in only one paper they published. The most dominant author, that is the author who is mostly the first author in multi-authored articles, is Peter Jones, with 3 first-authored articles.

4.2. Intellectual structure of the field

To begin analyzing the intellectual structure of the field, we first bring the articles that were most cited by our database (Table 5). As we can see, the most cited article is a methodological one, which is mostly cited in papers that use structural equations as a method (Fornell & Larcker, 1981). This suggests that research in retail and privacy was mostly done by this kind of method. Differently, the fifth most cited paper is a theoretical one,

that deals with relationship marketing (Morgan & Hunt, 1994), an important theory that is discussed in privacy and retail, as we can see for the first and second articles with the most citations in our database, Garbarino & Strahilevitz (2004) and Eastlick et al. (2006), respectively.

Table 5. Most cited articles by the database

CT	Title	Authors	Year	Source
24	Evaluating Structural Equation Models with Unobservable Variables and Measurement Error	Claes Fornell; David F. Larcker	1981	Journal of Marketing Research
16	A secure and privacy-preserving mobile wallet with outsourced verification in cloud computing	Zhen Qin; Jianfei Sun; Abubaker Wahaballa; Wentao Zheng; Hu Xiongb; Zhiguang Qin	2017	Computer Standards & Interfaces
12	Measuring Service Quality in E-Retailing	Joel E.Collier; Carol C.Bienstock	2006	Journal of Service Research
10	Consumer Perceptions of Privacy and Security Risks for Online Shopping	Anthony D. Miyazaki; Ana Fernandez	2001	Journal of Consumer Affairs
10	The Commitment-Trust Theory of Relationship Marketing	Robert M. Morgan; Shelby D. Hunt	1994	Journal of Marketing

Notes: CT = number of times cited in our database
Source: Elaborated by the authors (2022)

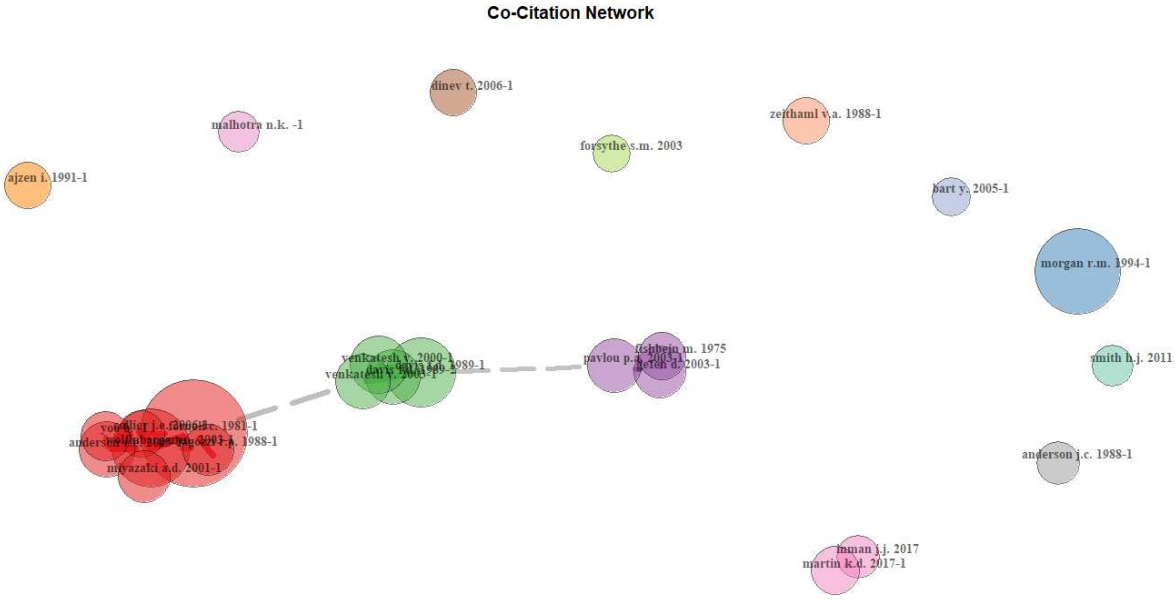
Another way to look at the intellectual structure is by using a co-citation network (Figure 3). The co-citation network is a way to visualize how articles have been cited within each other. The full co-citation of our database has a size of 18,638 vertices and a density of 0.005. This indicates that although there are a lot of citations in our database, they don't relate too much to each other. In other words, the intellectual structure of the field is fragmented.

In Figure 3, we plotted the co-citations network from our main 25 cited references. We can observe three clusters that relate to each other. The biggest cluster (the red one in our Figure) is composed of articles more related to a method and measuring, containing the most cited article Fornell & Larcker (1981) and related ones like Wolfinbarger & Gilly (2003) and Bagozzi & Yi (1988). But also in this cluster are articles that applied regression models in retail like Anderson & Srinivasan (2003) and Miyazaki & Fernandez (2001). The second cluster is composed of 4 articles and is more related to technology acceptance containing the seminal article by Davis (1989) and others that discuss the Technology Acceptance Model, like Venkatesh, Morris, Davis and Davis (2003). These two clusters are highly related as we can see by the gray dashed lines. The other related cluster is one composed of three articles that deal with consumer behavior and the application of TAM, like Gefen, Karahanna, and Straub (2003) that integrated the trust and the TAM perspectives in online shopping.

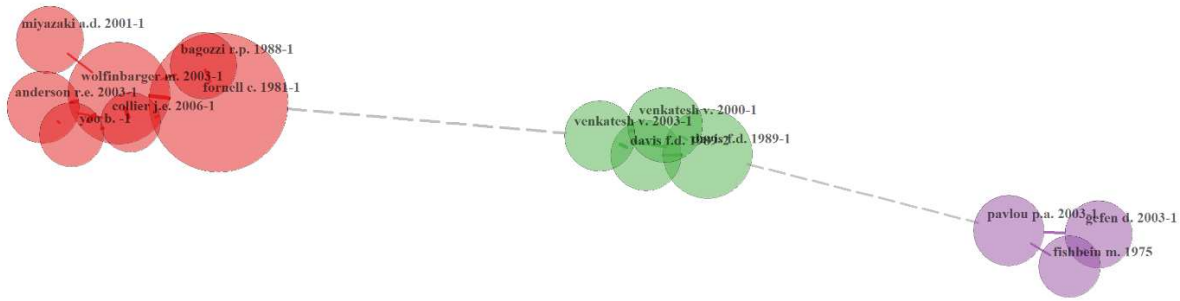
The main intellectual structure of the field is composed mainly of seminal articles that serve as theoretical or methodological bases. One exception is the cluster composed by Inman & Nikolova (2017) and Martin, Borah, and Palmatier (2017). These articles deal with how modern technology affects consumers and firms, and how to deal with consumer data vulnerability perceptions. The fact that, from the 25 most cited articles by our database, there are two from 2017 shows the recent relevance of the theme and how the theoretical background is keeping up with the pace.

Figure 3. Co-citation network from the 25 most cited articles from the database

Panel A: Full co-citation network



Panel B: Zoom of the three main clusters of the co-citation network



Source: Elaborated by the authors (2022)

4.3. Conceptual structure

To have a more accurate view of the field, we conducted a conceptual structure analysis. The conceptual structure is based on the keywords that are most used in our database and how they relate to each other. Table 6 shows the ten most used keywords and Figure 4 plots the relationship between the top thirty articles keywords-plus.

Looking at the most used keywords, we can see that electronic commerce is one theme that is heavily linked with privacy and retail. We can also see a keyword of a specific technology that is radio frequency identification (RFID) that was very used in our database. Although not new, RFID is a technology being used nowadays to monitor supply chains and it is heavily related to retail and technology, as it may increase return over investments (ROI) for retailers (Attaran, 2007).

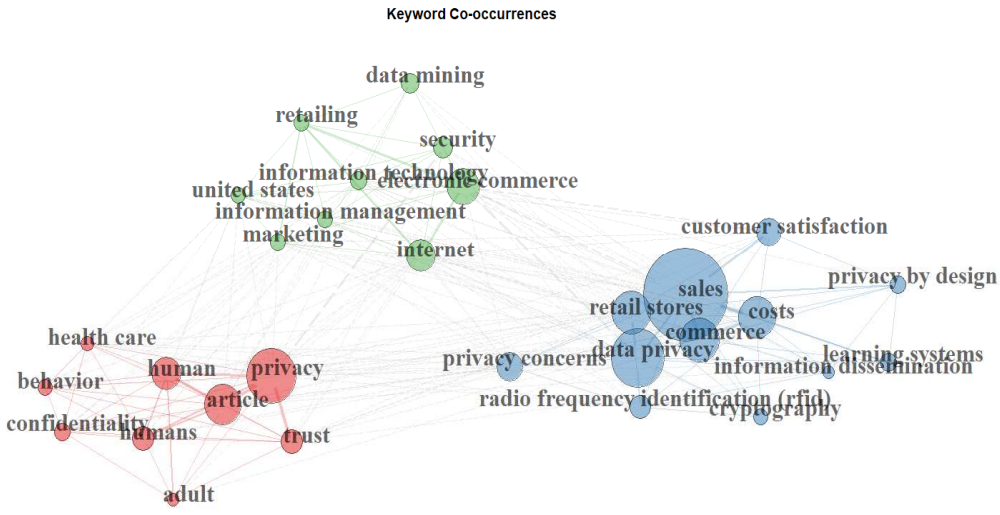
Table 6. Most Relevant Keywords

Rank	Author Keywords	Articles	Articles Keywords-Plus	Articles
1	PRIVACY	57	PRIVACY	47
2	TRUST	29	SALES	44
3	RETAILING	28	DATA PRIVACY	31
4	RFID	24	ELECTRONIC COMMERCE	28
5	E-COMMERCE	22	RETAILING	28
6	SECURITY	19	INTERNET	25
7	ELECTRONIC COMMERCE	15	RETAIL STORES	21
8	ETHICS	14	COMMERCE	18
9	ONLINE RETAILING	14	COSTS	18
10	RETAIL	14	TRUST	18

Notes: Author Keywords = keyword mentioned by the authors in their articles; Articles Keywords-Plus = keywords generated by the scientific databases (Scopus and WoS)
Source: Elaborated by the authors (2022)

Observing Figure 4 we can see three clusters. These clusters are the main conceptual structure of the field of retail and privacy. In the biggest cluster (the blue one), the keywords are more concerning to firms, like sales, costs, and information dissemination. This was expected since research in retail tends to be very applied so it can be useful to managers. The other two clusters are similar in size, one dealing more with issues related to consumer behavior in general (the red one) and the other dealing with issues of the online environment, with keywords like electronic commerce, data mining, and information technology (the green one).

Figure 4. Network of keywords

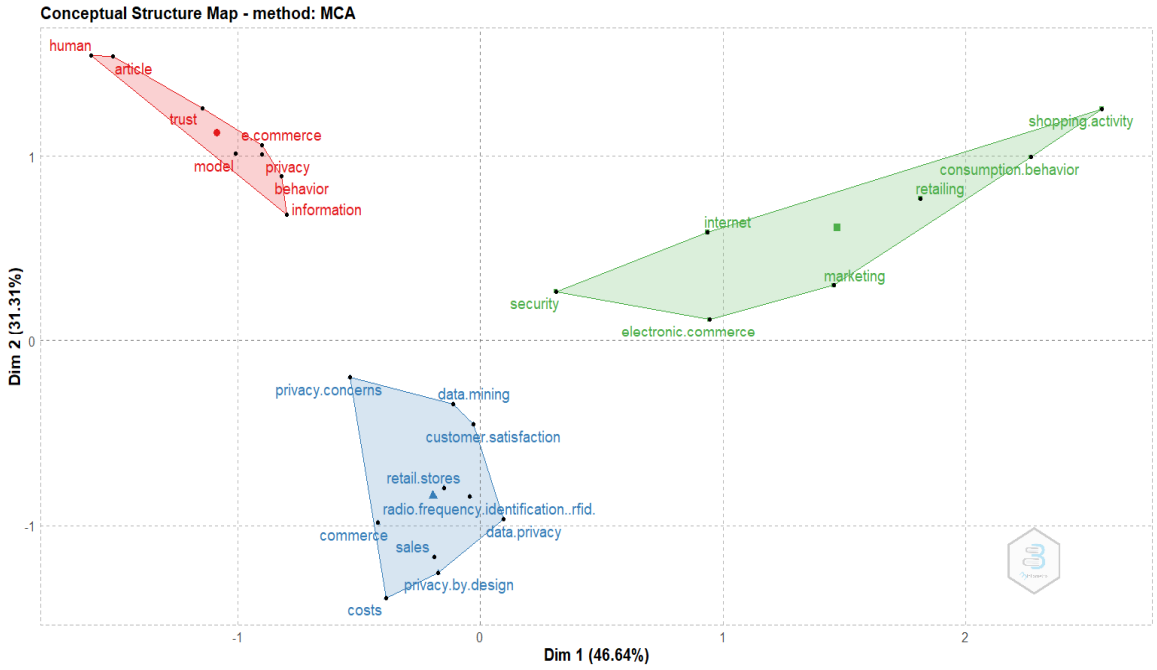


Source: Elaborated by the authors (2022)

In Figure 5, we use the multiple correspondence analysis (MCA) method to visualize the field by keywords in a two-dimensional plot. MCA is a technique for analyzing multivariate categorical data and visualizing this data in a two-dimensional plot (Greenacre & Blasius,

2006). MCA allows us to see the proximity between the keywords in their respective clusters. For example, we can see that although the cluster dealing with consumer behavior and the cluster of the online environment are similar in size, the first one is much more aggregated than the second. This means that articles in the behavior cluster are more alike than the articles in the online environment cluster.

Figure 5. Conceptual structure map of the bibliographic database



Source: Elaborated by the authors (2022)

In the next section, we will discuss the main articles in each cluster and propose a research agenda focused on the clusters.

5. Discussion and Research Agenda

We will start analyzing the clusters by their size. First, we will discuss the cluster that has keywords more related to the firm. Then, the second cluster of consumer behavior-related keywords will be discussed. Last, the cluster that is related to the online environment. Although we used this nomenclature to aggregate the keywords in clusters, as it will be seen, the online characteristic is present in basically all articles in all clusters.

5.1. Data Privacy and the firm

As seen by the analysis of keywords in the previous section, this cluster is composed mainly of articles that deal with business performance. Another characteristic of this cluster is that it is mostly composed of articles from the 2000's decade, which shows that in that decade the concern of marketing researches regarding retail and privacy was more about firm performance.

In this cluster are present articles like Kurkovsky & Harihar (2006) that introduced a brand-new technology for their time that was a mobile marketing application, which could deliver targeted promotions to the customer's mobile phone. They proposed that this application (named SMMART by them) would keep information about consumer security and

privacy was one of their main concerns. In addition, they showed that this application could increase sales revenue for retailers, by conducting experiments based on the implementation costs and future sales. Even 10 years later, this integration between mobile devices and consumer non-online shopping was the focus of a research agenda proposed by Kannan & Li (2017).

Another example from this cluster is Bhattacharjee, Gopal, Lertwachara, and Marsden (2006) which analyzed online music sharing and its implications for retailers and consumers. They dealt with consumer search costs in the legal environment (brick and mortar cd stores, online music stores) and the illegal online environment. They proposed that the low quality of the music in the illegal environment could be explored by the legal music retailers and increase their profits. Interestingly, this subject is also discussed by Kannan and Li (2017), in how digital marketing could help to decrease piracy through digital rights management (DRM). Other articles present in this cluster are Smith, Szongott, Henne, and von Voigt (2012) which discuss big data privacy in public social media, and Chiu et al. (2009) which discuss an extension of the Technology Acceptance Model to analyze consumer online repurchase behavior, by introducing trust and enjoyment along with the ease of use and usefulness as antecedents of repurchase and how this benefits the firm. Based on what was discussed and what could be relevant for marketing research about privacy and retail and firm performance, we propose the following research topics:

i. Artificial Intelligence (AI) and the Internet of Things (IoT) are becoming more present each day in consumers' lives. It is known that problems related to privacy negatively affect the firm, and consumers value privacy. Research could measure what impact the privacy-related issues specifically related to AI and IoT will have on retail performance in the long term.

ii. Although piracy is heavily related to privacy loss, there is still a lot of piracy in the online environment. Future research could analyze what online retail stores could do to reduce piracy based on consumers' privacy concerns.

iii. We realized the great relevance of the IT area for data collection and analysis, especially for thinking about data privacy. So, we propose, as a research question, how could marketing and IT departments work together to think about customer data management, so that personalization could be offered without being intrusive?

5.2. Data privacy and consumer behavior

Another cluster in our database is more related to words concerning consumer behavior, such as trust and privacy. Although consumer behavior is a theme that has been studied for a long time, the discussion of consumer behavior related to data privacy seems to be up to date, since there are articles in this cluster from the last five years.

In this cluster, there are articles like Faraoni, Rialti, Zollo, and Pellicelli (2019) that discuss the role of security, privacy, and website design in consumer loyalty, especially applied to an online grocery retail context. One interesting finding of them and a possible avenue for future research was the high correlation between trust and website design (Faraoni et al., 2019).

Another example of research being done is Bui, Kettinger, and Park (2021) which analyzed adaptive personalization systems, systems that use the information provided by the consumer to provide customized solutions. On the other hand, those systems require, obviously, that the consumer shares their identity and information with the provider. They focused on consumers only that used online retail with adaptive personalization systems and

showed that willingness to share identity influenced repurchase intention through perceived personalization value.

As seen in our theoretical background, consumer trust is one of the main foundations for the success not only of retailing but for marketing at all (Dekimpe, 2020) and articles regarding privacy and retail in this area are dealing with new technologies and how consumers can trust them. Following this and discussions made by Cui et al. (2021), we propose the following research topics for consumer behavior and privacy for retail:

i. It is known that consumers share their data in exchange for benefits. So future research could focus on what kind of benefits consumers want for their data. Should these benefits be more utilitarian (like discounts)? Should they be more hedonic (like personalized experiences)? Is there an optimal point of shared data by consumers and benefits conceived by the firm to them?

ii. Thinking about hedonic and utilitarian purchases, future research should verify the different impacts of privacy for different product categories. For example, for more utilitarian purchases, with low involvement, would consumers be less concerned with their privacy?

5.3. Online environment and data privacy

As can be seen, by the keywords in the MCA analysis (Figure 5), this third cluster is very similar to the consumer behavior cluster. This was expected since research about the consumer is one of the main concerns of research in retail (Verhoef, 2021). However, articles in this cluster differentiate themselves as they focus more on the adoption of technology, like de Kerviler, Demoulin, and Zidda (2016) who investigated the role of financial and privacy risks as key drivers of in-store mobile payment.

Another example is the work by Faqih (2016) which investigated perceived ease-of-use, perceived usefulness, perceived compatibility, social influence, trust, perceived risk, privacy, security, Internet shopping anxiety, Internet self-efficacy, and price in the adoption of Internet shopping for a developing country. They found that perceived risk and privacy concerns do not directly influence behavioral intention to adopt internet shopping, but they are moderated by gender, as they influence the behavioral intention more of women.

In addition to that, there are articles like Poushneh (2018) that focus on one specific technology that is augmented reality (AR), and how to measure its quality. He focused on the augmentation quality, the quality of objects, and the quality of how virtual objects relate to the "real" world and shows that this quality is a driver of consumer satisfaction. Based on that and how firms can improve consumer online experience regarding retail and data privacy we propose the following research themes:

i. Future research should be on how to improve security sensation for consumers regarding their data privacy. Causal relations between retailer site features (like design) and consumer sensation of being safe could be sought. Since there is gender difference in privacy and security issues, could some features be more useful for reducing risk feelings for different genders?

Table 7 summarizes each of the clusters and the respective proposed future research agenda.

Table 7. Research agenda separated by clusters

Clusters	Research Agenda
Data Privacy and the firm	Artificial Intelligence (AI) and the Internet of Things (IoT) are becoming more present each day in consumers' lives. It is known that problems related to privacy negatively affect the firm, and consumers value privacy. Research could measure what impact the privacy-related issues specifically related to AI and IoT will have on retail performance in the long term.
	Although piracy is heavily related to privacy loss, there is still a lot of piracy in the online environment. Future research could analyze what online retail stores could do to reduce piracy based on consumers' privacy concerns.
	We realized the great relevance of the IT area for data collection and analysis, especially for thinking about data privacy. So, we propose, as a research question, how could marketing and IT departments work together to think about customer data management, so that personalization could be offered without being intrusive?
Data Privacy and consumer behavior	It is known that consumers share their data in exchange for benefits. So future research could focus on what kind of benefits consumers want for their data. Should these benefits be more utilitarian (like discounts)? Should they be more hedonic (like personalized experiences)? Is there an optimal point of shared data by consumers and benefits conceived by the firm to them?
	Thinking about hedonic and utilitarian purchases, future research should verify the different impacts of privacy for different product categories. For example, for more utilitarian purchases, with low involvement, would consumers be less concerned with their privacy?
Online Environment and data privacy	Future research should be on how to improve security sensation for consumers regarding their data privacy. Causal relations between retailer site features (like design) and consumer sensation of being safe could be sought. Since there is gender difference in privacy and security issues, could some features be more useful for reducing risk feelings for different genders?

6. Conclusions

Nineteen years ago, Parasuraman & Zinkhan (2002) said that privacy and security issues were already considered challenges for the future use of the Internet, but that call is apparently still unresolved. As technology evolves, data privacy issues become more severe. To confidently enjoy the benefits of technology, both managers and consumers require measures such as regulatory laws to protect personal data. Public policy makers are trying to develop regulations to make this technological environment more suitable in questions regarding to private data, like the creation of Brazil's LGPD (Lei Geral de Proteção de Dados Pessoais, nº 13.709/2018).

As the academia also tend to reflect the society concerns, we saw an increase of scientific publications regarding data privacy in retail, and a bibliometric review seemed appropriated in the stage that the field is. Therefore, the main goal of our research was to analyze the publications about privacy and retail. Our findings suggest that the citation structure of the field is mature, but fragmented. The main conceptual structure of the field, by keywords correspondence, showed three clusters of similar publications. One more related to firms, other to consumer behavior and another to preoccupations of the online environment. More than that, we proposed a research agenda based on the findings of the

literature review. Our main theoretical contribution is to identify the current state of academic production about data privacy and retail. Therefore, we expand retail literature by bringing data privacy to the center of a bibliometric analysis and by proposing a research agenda focused on this area of research.

As practical contributions, our research may help retail managers to understand the importance of customers' data privacy, as this subject is studied since 1975. In addition, diverse publications in our literature analysis suggest that if well managed, customer privacy can be a competitive advantage for firms. More than that, consumers are willing to trade some of their data to gain benefits. Managers that are wary of that, and who build trust relations with their customers tend to benefit from consumers' data privacy concerns.

However, it is important to take into account the limitations related to the study that demonstrate future paths. Our bibliometric analysis considered articles from two databases: Web of Science and Scopus. Even though these are the most commonly used databases, it would be interesting for future research to compare the results with other databases, such as Google Scholar, to better review the data. Another future path would be to advance from the bibliometric study to a structured literature review, understanding the main methods, theories, and constructs of the privacy literature. It is important to emphasize that theoretically based analyzes would imply a critical literature review approach, which is not the purpose of bibliometric studies such as ours (Martínez-López et al., 2018).

RIMAR Note

A preliminary version of this article was presented at BALAS Annual Conference 2022.

References

Anderson, R. E., & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology and Marketing,* 20(2), 123–138. https://doi.org/10.1002/mar.10063

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Attaran, M. (2007). RFID: An enabler of supply chain operations. *Supply Chain Management: An International Journal*, 12(4), 249–257. https://doi.org/10.1108/13598540710759763

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, 16, 74-94. https://doi.org/10.1007/BF02723327

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., & Marsden, J. R. (2006). Consumer Search and Retailer Strategies in the Presence of Online Music Sharing. *Journal of Management Information Systems*, 23(1), 129–159. https://doi.org/10.2753/MIS0742-1222230104

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing,* 93(1), 79–95. https://doi.org/10.1016/j.jretai.2016.12.004

Bui, S., Kettinger, W. J., & Park, I. (2021). Identity Sharing and Adaptive Personalization Influencing Online Repurchases. *Journal of Computer Information Systems*, 1–12. https://doi.org/10.1080/08874417.2021.1919939

Chiu, C., Chang, C., Cheng, H., & Fang, Y. (2009). Determinants of customer repurchase intention in online shopping. *Online Information Review*, 33(4), 761–784. https://doi.org/10.1108/14684520910985710

Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*. Retrieved from: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Cui, T. H., Ghose, A., Halaburda, H., Iyengar, R., Pauwels, K., Sriram, S., Tucker, C., & Venkataraman, S. (2021). Informational Challenges in Omnichannel Marketing: Remedies and Future Research. *Journal of Marketing*, 85(1), 103–120. https://doi.org/10.1177/0022242920968810

Culnan, M. J. (1995). Consumer awareness of name removal procedures: Implications for direct marketing. *Journal of Direct Marketing*, 9(2), 10–19. https://doi.org/10.1002/dir.4000090204

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319. https://doi.org/10.2307/249008

de Kerviler, G., Demoulin, N. T. M., & Zidda, P. (2016). Adoption of in-store mobile payment: Are perceived risk and convenience the only drivers? *Journal of Retailing and Consumer Services*, 31, 334–344. https://doi.org/10.1016/j.jretconser.2016.04.011

Dekimpe, M. G. (2020). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*, 37(1), 3–14. https://doi.org/10.1016/j.ijresmar.2019.09.001

Eastlick, M. A., Lotz, S. L., & Warrington, P. (2006). Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment. *Journal of Business Research*, 59(8), 877–886. https://doi.org/10.1016/j.jbusres.2006.02.006

Faqih, K. M. S. (2016). An empirical analysis of factors predicting the behavioral intention to adopt Internet shopping technology among non-shoppers in a developing country context: Does gender matter? *Journal of Retailing and Consumer Services*, 30, 140–164. https://doi.org/10.1016/j.jretconser.2016.01.016

Faraoni, M., Rialti, R., Zollo, L., & Pellicelli, A. C. (2019). Exploring e-Loyalty Antecedents in B2C e-Commerce: Empirical results from Italian grocery retailers. *British Food Journal*, 121(2), 574–589. https://doi.org/10.1108/BFJ-04-2018-0216

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research,* 18(1), 39. https://doi.org/10.2307/3151312

Garbarino, E., & Strahilevitz, M. (2004). Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. *Journal of Business Research*, 57(7), 768–775. https://doi.org/10.1016/S0148-2963(02)00363-6

Gefen, Karahanna, & Straub. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, 27(1), 51. https://doi.org/10.2307/30036519

Goswami, S. (2020, December 14). Council Post: The Rising Concern Around Consumer Data And Privacy. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2020/12/14/the-rising-concern-around-consumer-data-and-privacy/

Greenacre, M., & Blasius, J. (Eds.). (2006). *Multiple Correspondence Analysis and Related Methods* . Chapman and Hall/CRC. https://doi.org/10.1201/9781420011319

Grewal, D., Gauri, D. K., Roggeveen, A. L., & Sethuraman, R. (2021). Strategizing Retailing in the New Technology Era. *Journal of Retailing*, 97(1), 6–12. https://doi.org/10.1016/j.jretai.2021.02.004

Grewal, D., & Roggeveen, A. L. (2020). Understanding Retail Experiences and Customer Journey Management. *Journal of Retailing*, 96(1), 3–8. https://doi.org/10.1016/j.jretai.2020.02.002

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing*, 93(1), 1–6. https://doi.org/10.1016/j.jretai.2016.12.008

Guha, A., Grewal, D., Kopalle, P. K., Haenlein, M., Schneider, M. J., Jung, H., Moustafa, R., Hegde, D. R., & Hawkins, G. (2021). How artificial intelligence will affect the future of retailing. *Journal of Retailing*, 97(1), 28–41. https://doi.org/10.1016/j.jretai.2021.01.005

Inman, J. J., & Nikolova, H. (2017). Shopper-Facing Retail Technology: A Retailer Adoption Decision Framework Incorporating Shopper Attitudes and Privacy Concerns. *Journal of Retailing*, 93(1), 7–28. https://doi.org/10.1016/j.jretai.2016.12.006

Kannan, P. K., & Li, H. A. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22–45. https://doi.org/10.1016/j.ijresmar.2016.11.006

Kurkovsky, S., & Harihar, K. (2006). Using ubiquitous computing in interactive mobile marketing. *Personal and Ubiquitous Computing*, 10(4), 227–240. https://doi.org/10.1007/s00779-005-0044-5

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317–323.

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data Privacy: Effects on Customer and Firm Performance. *Journal of Marketing*, 81(1), 36–58. https://doi.org/10.1509/jm.15.0497

Martínez-López, F. J., Merigó, J. M., Valenzuela-Fernández, L., & Nicolás, C. (2018). Fifty years of the European Journal of Marketing: A bibliometric analysis. *European Journal of Marketing*, 52(1/2), 439–468. https://doi.org/10.1108/EJM-11-2017-0853

Meehan, M. (2019, November 16). Data Privacy Will Be The Most Important Issue In The Next Decade. *Forbes*. https://www.forbes.com/sites/marymeehan/2019/11/26/data-privacy-will-be-the-most-important-issue-in-the-next-decade/

Miyazaki, A. D., & Fernandez, A. (2001). Consumer Perceptions of Privacy and Security Risks for Online Shopping. *Journal of Consumer Affairs*, 35(1), 27–44. https://doi.org/10.1111/j.1745-6606.2001.tb00101.x

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, 58(3), 20. https://doi.org/10.2307/1252308

Parasuraman, A., & Zinkhan, G. M. (2002). Marketing to and Serving Customers through the Internet: An Overview and Research Agenda. *Journal of the Academy of Marketing Science*, 30(4), 286–295. https://doi.org/10.1177/009207002236906

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*, 19(1), 27–41. https://doi.org/10.1509/jppm.19.1.27.16941

Poushneh, A. (2018). Augmented reality in retail: A trade-off between user's control of access to personal information and augmentation quality. *Journal of Retailing and Consumer Services*, 41, 169–176. https://doi.org/10.1016/j.jretconser.2017.12.010

Roggeveen, A. L., & Sethuraman, R. (2018). Understanding the JR Heritage, Publishing in JR , and the Evolving Retail Field. *Journal of Retailing*, 94(1), 1–4. https://doi.org/10.1016/j.jretai.2018.02.001

Smith, M., Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media. 2012 *6th IEEE International Conference on Digital Ecosystems and Technologies (DEST),* 1–6. https://doi.org/10.1109/DEST.2012.6227909

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425. https://doi.org/10.2307/30036540

Verhoef, P. C. (2021). Omni-channel retailing: Some reflections. *Journal of Strategic Marketing*, 29(7), 608–616. https://doi.org/10.1080/0965254X.2021.1892163

Wolfinbarger, M., & Gilly, M. C. (2003). eTailQ: Dimensionalizing, measuring and predicting etail quality. *Journal of Retailing*, 79(3), 183–198. https://doi.org/10.1016/S0022-4359(03)00034-4

Zhou, Y., Lu, S., & Ding, M. (2020). Contour-as-Face Framework: A Method to Preserve Privacy and Perception. *Journal of Marketing Research*, 57(4), 617–639. https://doi.org/10.1177/0022243720920256

Authors

1. Alina Flores is Master in Managerial Sciences, Federal University of Rio Grande do Sul – UFRGS, and Ph.D. candidate at UFRGS.

2. Lucas Dorneles is Master in Managerial Sciences, Federal University of Rio Grande do Sul – UFRGS, and Ph.D. candidate at UFRGS.

Contribution of each author

Contribution	Alina Flores	Lucas Dorneles
1. Definition of research problem	✓	✓
2. Development of hypotheses or research question (in case of empirical studies)	✓	✓
3. Development of theretical propositions (in case of theoretical work)		
4. Theoretical foundation / literature review	✓	
5. Definition of methodologial procedures	✓	✓
6. Data collection		✓
7. Analysis and interpretation of data (when applied)	✓	✓
8. Revision of the manuscript	✓	
9. Manuscript writing	✓	✓