Vivat Academia. Revista de Comunicación (2024).
ISSN: 1575-2844

RESEARCH

Received: 26/04/2023 **Accepted**: 19/06/2023 **Published**: 11/09/2023

MILLENNIALS AND GENERATION X FACING THE REALITY OF BIG DATA AND THE PROTECTION OF PERSONAL DATA ON THE INTERNET

Millennials y Generación X frente a la realidad del Big data y la protección de datos personales en Internet

Inmaculada Martín Herrera[1]**:** University of Pablo de Olavide. Spain.
imartin@centrosanisidoro.es

Mercedes Guerrero Caballero: University of Pablo de Olavide. Spain.
mercedesguecab@gmail.com

How to cite this article:

Martín-Herrera, Inmaculada, & Guerrero-Caballero, Mercedes (2024).Millennials and Generation X Facing the Reality of Big Data and the Protection of Personal Data on the Internet [Millennials y Generación X frente a la realidad del Big data y la protección de datos personales en Internet]. *Vivat Academia*, 157, 1-20. http://doi.org/10.15178/va.2024.157.e1489

ABSTRACT

The telematic procedures that are developed in collaborative environments are nowadays daily processes *in crescendo*. Its execution implies the transfer of personal data, which is stored by the organizations involved, without, in most cases, the users of these services and platforms being able to control the processing and the uses made of them. In this context, the intention of this study is to investigate the perceptions of Spanish citizens of legal age regarding the management of their personal data in cyberspace and its relationship with big data. Likewise, it is intended to know how two personal variables influence these attitudes: gender and age, in order to determine if there is a gap between men and women, and between users of two population generations: Millennials and Generation X. The data was collected using a scale consisting of 20 items and administered to a sample of 344 people. On the other hand, the tests to analyze this information were a descriptive study and a dependency analysis, which has made it possible to detect discrepancies in the results based on age.

[1] **Inmaculada Martín Herrera**: Doctor in Education Sciences. Graduated in Communication. Professor and researcher at the San Isidoro University Center, affiliated with the Pablo de Olavide University (Spain).

The results conclude that users show meager knowledge about the concept and operation of big data; and they express, on the one hand, their interest in knowing how their data is used and, on the other, their concern that they may be involved in problems related to their security and the commission of cybercrimes.

Keywords: Big data, Internet, data protection, security, user, algorithm, digitization.

RESUMEN

Las gestiones telemáticas que se desarrollan en entornos colaborativos constituyen hoy en día procesos cotidianos *in crescendo*. Su ejecución implica la cesión de datos personales, que son almacenados por los organismos implicados, sin que, en la mayoría de los casos, los usuarios de estos servicios y plataformas puedan controlar el procesamiento y los usos que se hace de los mismos. En este contexto, la intención del presente estudio es indagar en las percepciones de la ciudadanía española mayor de edad en torno a la gestión de sus datos personales en el ciberespacio y su correlación con el *big data*. Igualmente, se pretende conocer cómo influyen en estas actitudes dos variables personales: el género y la edad, con objeto de determinar si existe una brecha entre hombres y mujeres, y entre los usuarios de dos generaciones poblacionales: los Millennials y la Generación X. Los datos fueron recabados mediante una escala que consta de 20 ítems y que se administró a una muestra integrada por 344 personas. Por su parte, las pruebas para analizar esta información fueron un estudio descriptivo y un análisis de dependencia, que han permitido detectar discrepancias en los resultados en función de la edad. Los resultados concluyen que los usuarios demuestran un conocimiento exiguo sobre el concepto y el funcionamiento del *big data*; y manifiestan, por un lado, su interés por saber cómo se emplean sus datos y, por otro, su preocupación por verse implicados en problemas relacionados con su seguridad y con la comisión de ciberdelitos.

Palabras clave: *Big data*, Internet, protección de datos, seguridad, usuario, algoritmo, digitalización.

1. INTRODUCTION

The global village that the world has become experiences high connectivity among people, and this activity results in a vast amount of information that circulates permanently through numerous channels. It is estimated that the planetary volume of this data will increase by 530% by 2025 compared to 2018, and it is a crucial factor in digital transformation (European Commission, 2020).

This productivity has led to the existence of massive data that requires systematic management, a process that has given rise to the concept known as big data, data intelligence, or macrodata. Laney (2001) was one of the first authors to refer to this phenomenon, which he defines as information assets with a large volume, variety, and velocity that require innovative and cost-effective ways of processing information in order to enhance knowledge and decision-making. In turn, this conglomerate of content with personal user information, expressed through different formats, originating from diverse sources, circulating and generating at high speed, becomes

an opportunity for organizations to discover hidden business patterns, consumer preferences, and market trends.

Soto (2017) points out that, in many cases, little or nothing is known about the subsequent uses and treatment of data in cyberspace by external companies that buy or rent such data for unauthorized purposes from their owners, thus violating fundamental rights such as privacy, confidentiality, and autonomy.

At this point, it is necessary to question to what extent citizens are aware of the destination and employability of the data they have voluntarily provided and how they self-protect their right to privacy on the Internet. Taking this research problem as a reference, the study hypotheses are proposed: h1) Despite the information available to them, users lack the necessary knowledge and do not pay enough attention to the transfer of their data in cyberspace; h2) There are differences in the citizens' perception of big data, the management of their personal data on the Internet, and their privacy based on the gender variable; h3) There are differences in this citizens' perception based on the age variable, so that the generation to which individuals belong, Millennials or Generation X, conditions their attitudes and behaviors.

1.1. Privacy and Personal Data Management on the Internet

Privacy, according to the Royal Spanish Academy (n.d.), is the "sphere of private life that one has the right to protect from any intrusion." If we extend this concept to digital reality, in addition to being any key content, in textual or audiovisual format, that allows identifying or tracking a person, personal data linked to an individual's privacy also includes information that allows tracking their behavior on the Internet. Therefore, the General Data Protection Regulation (European Parliament, 2018) also interprets as personal data other elements such as cookies, mobile phone identifiers, or IP addresses, and warns about those that are especially sensitive, such as religion or those related to medical history.

According to González (2015), the use of technology can generate damages to people's privacy that are very difficult or impossible to repair, either due to not taking the necessary precautions or through malicious actions by third parties. In this context, two fundamental factors come into play: on one hand, the knowledge and common sense of the user; and on the other hand, the Law. Regarding the legal framework, digital privacy in Spain is regulated by the Organic Law 3/2018, of December 5, on the Protection of Personal Data and Guarantee of Digital Rights (LOPDGDD) (2018), which adapts the General Data Protection Regulation (GDPR) that applies at the European level to Spanish legislation.

Meraz-Espinoza (2018) points out that in the context of the Internet, privacy means informational autonomy, understood as the right of individuals to decide what personal information will be disclosed, to whom, and for what purposes, with the intention of having control over this data and protecting themselves from external threats, including doxing, phishing, and identity theft.

Faced with these dangers, authors and organizations (Spanish Data Protection Agency, 2018; Suras, 2020; Zúñiga-Becerra, 2018) propose strategies aimed at safeguarding privacy and protecting the personal data we handle when operating in cyberspace. Some of them are as follows: properly configure the privacy settings of digital applications; avoid sharing passwords and sensitive information on social networks; protect the digital footprint; create strong and different passwords for each network or system; install antivirus software; do not accept or block suspicious users; and read the terms and conditions of use before accessing an online platform (Figure 1).

Figure 1

Risks and Privacy Protection Strategies on the Internet.

Source: Author's own work.

Regarding big data, Escobar and Mercado (2019) consider that the globalization of information and mobile connection systems has accelerated and facilitated a massive and daily transmission of digital information. These data are incorporated into technological platforms, favoring and streamlining various everyday processes in diverse fields such as finance, education, and healthcare. Moreover, they represent an advancement for society and have improved living conditions. Examples of this include GPS devices or the tool Google Maps, as well as smart energy supply in cities, public safety, and personalized advertising with more accurate recommendations.

The described situation has led to the need to optimize storage models and the processing of this information. Both realities converge in the phenomenon known as big data, a concept linked to the technologies used in managing massive volumes of data derived from various sources and generated rapidly (Hernández-Leal et al., 2017). The term big data refers to the constant collection, analysis, and accumulation of large amounts of data, including personal data, from different sources and subject to automated processing using computer algorithms and advanced processing techniques. This includes both stored and continuously transmitted data, with the aim of generating correlations, trends, and patterns (European Parliament, 2017).

2. OBJECTIVES

It is of interest to investigate the perceptions of adult internet users regarding the management of their personal data in cyberspace and, in correlation with this topic, their knowledge about the concept and functioning of big data. For this purpose, the following objectives are proposed in this research: a) Determine the opinions of a sample of the Spanish population on issues such as privacy on the Internet, the handling of their personal information by third parties, and the role that big data plays in this reality, b) Determine how two personal variables, gender and age (specifically, two age groups corresponding to two generations: Millennials and Generation X), influence these perceptions, c) Understand the relationship between these variables and the users' perceptions to establish, if possible, an explanatory link between them.

3. METHODOLOGY

The study follows a descriptive methodology with a quantitative approach, where data was collected through a self-developed measurement instrument: a semi-structured questionnaire. This questionnaire consists of 20 items with dichotomous (yes or no), multiple-choice, open-ended, and value-scale responses. The value scale ranges from 1, meaning "Not at all in agreement (NA)," to 4, meaning "Completely in agreement (TA)." The questions have been categorized into two thematic blocks for the analysis of the results: 'Privacy and Management of Personal Data on the Internet' (items 01 to 14) and 'Knowledge of Big Data' (items 15 to 20).

In addition to this, there is a section related to the personal data of the research participants, corresponding to the studied variables: gender and age. Specifically, two age groups are considered, ranging from 19 to 39 years old and from 40 to 56 years old, which represent the differences between two generations, Millennials, and Generation X, respectively (Rubio-Laborda et al., 2021). The intention is to conduct a comparative study between these two generations, as they access the internet differently, both in terms of usage patterns and the quantity and frequency of use. Millennials are individuals who were born into the digital society, and they have been educated and working with communication technologies from the beginning, incorporating them into their daily tasks and projects. On the other hand, Generation X had to adapt to the digital revolution. With respect to these groups, the study is of a non-experimental type, as variables are not deliberately manipulated, no specific situation is constructed, the independent variables have already occurred and their effects are not altered, as they have already happened (Hernández-Sampieri et al., 2006).

The sample consists of 344 individuals, all adults, residing in Spain, who are regular internet users. The sample includes both men (47.1%) and women (52.6%), and their ages range from 19 to 56 years old. Regarding the age variable, the described sample has a nearly equal representation of subjects belonging to Generation X and Millennials, each accounting for about 50% of the total.

The data analysis was performed using the SPSS software, and two types of tests were conducted: a descriptive study at the item level with frequencies and percentages, and a dependence analysis using the calculation of Pearson's chi-square correlation

coefficient (X2). The purpose of the chi-square test was to determine if the differences obtained by the studied groups based on the variables are statistically significant. When the significance level (p) is equal to or less than 0.05, the variable is considered significant, leading to the rejection of the null hypothesis and indicating that there are differences between the groups in that particular variable.

The fieldwork was conducted during the year 2022. To carry it out, the questionnaire was designed in digital format using Google Forms and was distributed through various channels: email, the social media platform Instagram, and a QR code (Figure 2) that was shared through a TikTok channel.

Figure 2

Questionnaire access QR code.

Source: Author's own work.

Before its administration, a pretest of the scale was conducted with the participation of seven anonymous users. Their feedback helped to improve the instrument. Some of their comments were as follows: changing the order of questions 4 and 5 (previously in positions 11 and 12), adding response options 'to conduct a survey' and 'to participate in a giveaway' to item 14 ('Occasions when I have provided false information'), and replacing the dichotomous response (yes or no), which is more polarized, with a value-scale in question 15: 'I know how big data works.'

4. RESULTS

In this section, we proceed to present the results of the study, which address the pursued objectives, based on the two dimensions studied: 'Privacy and Management of Personal Data on the Internet' and 'Knowledge of Big Data'. In relation to the first dimension (Tables 1.A and 1.B), almost all individuals in the sample (95.9%) use the internet to a considerable or significant extent (item 1), and this usage requirement or factor has facilitated their active and effective participation in the study. The dependence analysis reveals significant differences in responses based on the age variable. Specifically, younger individuals or Millennials (X2 = 15.11, p = 0.002) make

greater use of the internet (90.7% marked the response option 'Completely in agreement') compared to those belonging to Generation X (77.3%).

Table 1.A
Results of the study in the dimension of Privacy and Management of Personal Data on the Internet.

Response options.	Total	Age		Gender	
		Millennials	Generation X	Man	Woman
I use the Internet.					
Not at all in agreement	0,6%	0,0%	1,2%	0,0%	1,1%
Slightly in agreement	3,5%	0,6%	6,4%	1,8%	5,0%
Fairly in agreement	11,9%	8,7%	15,1%	12,9%	11,0%
Completely in agreement	84,0%	90,7%	77,3%	85,3%	82,9%
Regarding my personal data on the Internet and the sharing of this information with others.					
I provide my data in some circumstances.	38,1%	37,2%	39,0%	36,2%	39,8%
I am indifferent about sharing my personal data.	14,5%	26,2%	2,9%	16,0%	13,3%
I do not share my data with others.	36,9%	19,8%	54,1%	41,1%	33,1%
I have never thought about this issue.	10,5%	16,9%	4,1%	6,7%	13,8%
I am willing to share my data in exchange for:					
Promotions and free products	37,8%	45,3%	30,2%	39,3%	36,5%
Free samples	32,0%	34,3%	29,7%	32,5%	31,5%
Discount vouchers	35,2%	35,5%	34,9%	35,6%	34,8%
Participating in raffles/draws	31,7%	31,4%	32,0%	35,0%	28,7%
Obtaining a subscription	29,1%	31,4%	26,7%	30,7%	27,6%
Never, I am not willing to share my data	39,8%	32,6%	47,1%	38,0%	41,4%
Other	3,8%	4,1%	3,5%	3,1%	4,4%
What data would you be willing to share with third parties in exchange for compensation?					
None	24,7%	13,4%	36,0%	23,9%	25,4%
E-mail	61,3%	73,3%	49,4%	60,1%	62,4%
Name	61,9%	70,9%	52,9%	60,1%	63,5%
Surnames	35,2%	50,6%	19,8%	35,6%	34,8%
Age	48,5%	65,7%	31,4%	48,5%	48,6%
Gender	45,3%	65,7%	25,0%	44,8%	45,9%
Telephone number	17,4%	32,0%	2,9%	17,2%	17,7%
National Identity Card	10,2%	20,3%	0,0%	9,8%	10,5%
Address	10,5%	18,0%	2,9%	11,7%	9,4%
I am concerned that others may obtain my personal data.					
Not at all in agreement	9,0%	16,3%	1,7%	9,8%	8,3%
Slightly in agreement	13,7%	22,7%	4,7%	16,6%	11,0%
Fairly in agreement	22,4%	34,9%	9,9%	17,2%	27,1%
Completely in agreement	54,9%	26,2%	83,7%	56,4%	53,6%
I am aware of how others handle my data, for example, how they store it or how they protect it.					
Not at all in agreement	60,5%	47,7%	73,3%	58,9%	61,9%
Slightly in agreement	24,7%	32,0%	17,4%	26,4%	23,2%
Fairly in agreement	9,9%	12,8%	7,0%	9,8%	9,9%
Completely in agreement	4,9%	7,6%	2,3%	4,9%	5,0%
I am aware of the use that others make of my data, whether they use it for commercial purposes or if they sell it to third parties.					
Not at all in agreement	58,1%	45,9%	70,3%	56,4%	59,7%
Slightly in agreement	23,8%	30,2%	17,4%	23,9%	23,8%
Fairly in agreement	11,0%	16,3%	5,8%	11,0%	11,0%
Completely in agreement	7,0%	7,6%	6,4%	8,6%	5,5%
I am concerned and interested in knowing the use that others may make of my personal data.					
Not all in agreement	9,0%	15,7%	2,3%	9,8%	8,3%
Slightly in agreement	11,6%	20,3%	2,9%	12,9%	10,5%
Fairly in agreement	16,9%	25,0%	8,7%	17,2%	16,6%

Response options.	Total	Age		Gender	
		Millennials	Generation X	Man	Woman
Completely in agreement	62,5%	39,0%	86,0%	60,1%	64,6%

Source: Author's own work.

Table 1.B

Results of the study in the dimension of Privacy and Management of Personal Data on the Internet.

Response options	Total	Age		Gender	
		Millennials	Generation X	Man	Woman
Regarding the previous question, what concerns you about the privacy of your data?					
Concerns about the privacy of my data include the disclosure of personal aspects.	54,9%	56,7%	53,2%	54,7%	55,1%
That others have access to my data.	62,4%	54,3%	70,2%	62,9%	61,9%
That my data may be stolen for committing crimes.	81,5%	81,1%	81,9%	84,3%	79,0%
That my data may be used to impersonate me.	57,6%	65,9%	49,7%	58,5%	56,8%
Other	2,1%	3,0%	1,2%	2,5%	1,7%
I take actions to control and know the treatment and use of my data.					
No	40,4%	50,0%	30,8%	36,8%	43,6%
Yes	59,6%	50,0%	69,2%	63,2%	56,4%
If you have selected "Yes" in the previous response, please indicate the actions you take:					
Reading privacy policies.	44,7%	43,7%	45,5%	39,8%	49,5%
Deleting personal information.	52,9%	43,7%	59,5%	55,3%	50,5%
Configuring the privacy settings of the profile.	63,5%	71,3%	57,9%	62,1%	64,8%
Disabling location tracking.	66,3%	67,8%	65,3%	69,9%	62,9%
Checking that it is a secure URL.	67,3%	62,1%	71,1%	73,8%	61,0%
Tracking suspicious sources.	17,8%	23,0%	14,0%	23,3%	12,4%
Other	2,4%	4,6%	0,8%	3,9%	1,0%
On some occasions, I have provided false information instead of my real data.					
No	34,6%	29,7%	39,5%	31,3%	37,6%
Yes	65,4%	70,3%	60,5%	68,7%	62,4%
If you have selected "Yes" in the previous response, on what occasions did you provide false information?					
Profile Creation	61,5%	74,4%	46,7%	57,1%	65,8%
For a subscription.	51,8%	40,5%	64,8%	58,9%	44,7%
To obtain discounts or free samples.	54,9%	40,5%	71,4%	60,7%	49,1%
To receive information by email.	54,9%	35,5%	77,1%	59,8%	50,0%
To complete a survey.	42,9%	31,4%	56,2%	48,2%	37,7%
To make online purchases.	29,6%	16,5%	44,8%	28,6%	30,7%
To participate in a raffle/draw.	16,4%	19,8%	12,4%	19,6%	13,2%
Other	2,2%	1,7%	2,9%	2,7%	1,8%

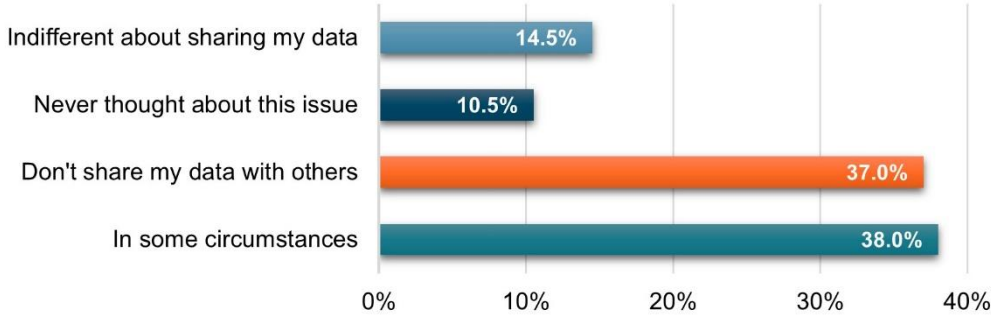
Source: Author's own work.

When asked about sharing their data with third parties (Figure 3), whether they are private companies, public administrations, search engines, or social networks (item 2), 14.5% of users are indifferent about providing them, 10.5% had not thought about this issue, 37% of respondents state that they avoid providing them, while 38% express that they share them on certain occasions. In this latter case, an open-ended question allowed us to know the circumstances (item 3). Thus, out of the 95 responses obtained, users indicate that they willingly and confidently share their data on websites of public administrations (29.5%), for online purchases (15.8%), when there is no other

alternative to carry out a procedure (13.6%), only in cases where it is a known and reliable platform (11.5%), for creating profiles on social networks or downloading a mobile application (7.4%), for work or academic purposes (6.6%), on bank websites (4.2%), for opening an email account (2.1%), on gaming platforms (2.1%), or for other reasons (7.2%), such as their interest in offers or promotions, participating in contests, subscribing to newsletters, and accessing digital media.

Figure 3

Responses to the item 'Regarding providing my data to third parties'.

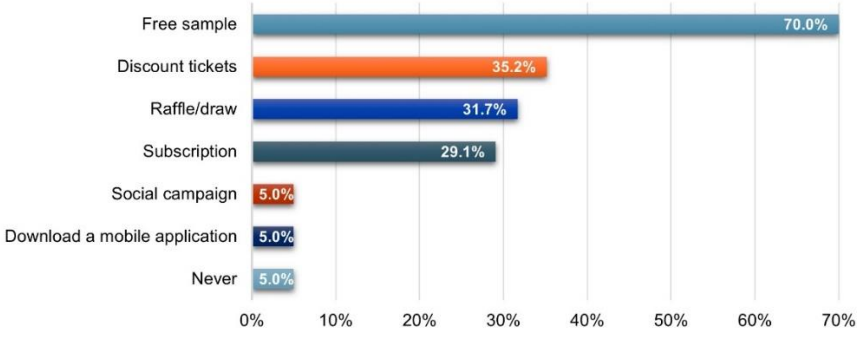


Source: Author's own work.

It is evident that users are willing to provide their data in exchange for some form of compensation (item 4) (Figure 4), such as free products (37.8%), samples (32.1%), discount vouchers (35.2%), participation in a raffle/draw (31.7%), or a subscription (29.1%); and to a much lesser extent, other options are mentioned, such as participating in a social advocacy campaign or downloading a mobile application. This result corresponds to the search for utility, benefit, profit, or interest that a product, service, or activity accessed through the internet provides to the user, whether they are a digital native or immigrant, thus satisfying their needs or desires (Hassan-Montero, 2006; Matellanes-Lazo, 2012). However, it is worth noting that nearly 40% indicate that they would not share their data for anything in return.

Figure 4

Responses to the item 'I would be willing to share my data in exchange for compensation'.



Source: Author's own work.

In these cases, regarding the type of data they would share (item 5), more than half of the users would provide their name (61.9%) or email address (61.3%). Meanwhile, 48.5% would disclose their age, and 45.3% would share their gender. However, these percentages decrease when it comes to sharing their phone number (17.4%), home address (10.5%), or identification document (10.2%).

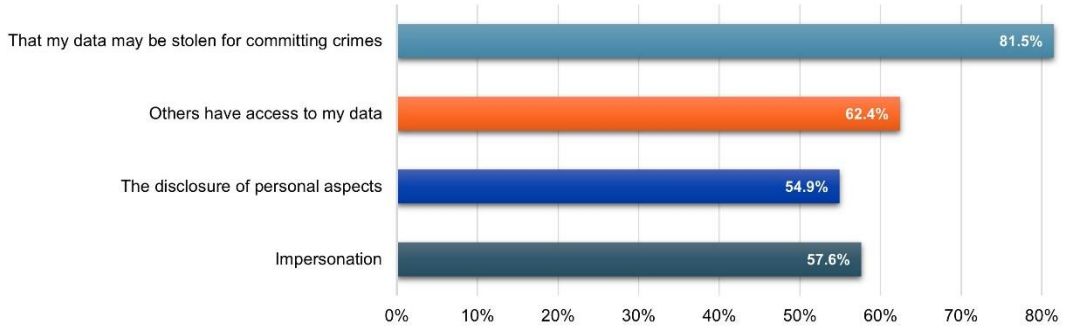
Regarding their concern about third parties obtaining and managing their data, the sum of affirmative responses confirms that the majority of the sample (77.3%) is quite or very concerned about this premise. This result can be correlated with the fact that more than 85% of the surveyed users claim to be unaware of how others treat their data, i.e., how they store or protect it. Likewise, a high percentage (81.9%) indicates that they do not know how these external agents use their personal information, whether they use it for commercial purposes or sell it to other organizations. In this regard, nearly 80% express being quite or very interested in knowing these uses.

The aspects that most concern users regarding their data on the Internet (item 10) (Figure 5) are data theft (81.5%), whether for fraud or committing crimes in their name; unauthorized access by others (62.4%); their internet browsing revealing private matters (54.9%); and identity theft (57.6%). In the 'Others' option, respondents have indicated concerns about protecting their rights to honor and privacy, information that may harm them professionally, or being bombarded with advertisements.

Considering the available statistics, it is understandable that citizens express their concerns regarding cybercrime. In this regard, a study by Cerezo-Domínguez and García-Cornejo (2020) highlights the increase in this type of crime over the last decade, with the most common being computer fraud and threats and coercion. The authors point out that these cybercrimes will continue to rise due to the digital transformation of both public and private organizations, the increasing population using ICT for communication, the challenges in prosecuting criminals due to issues related to extraterritoriality, and the perception of impunity.

Figure 5

Responses to the item 'Aspects that most concern users regarding the privacy of their data'.



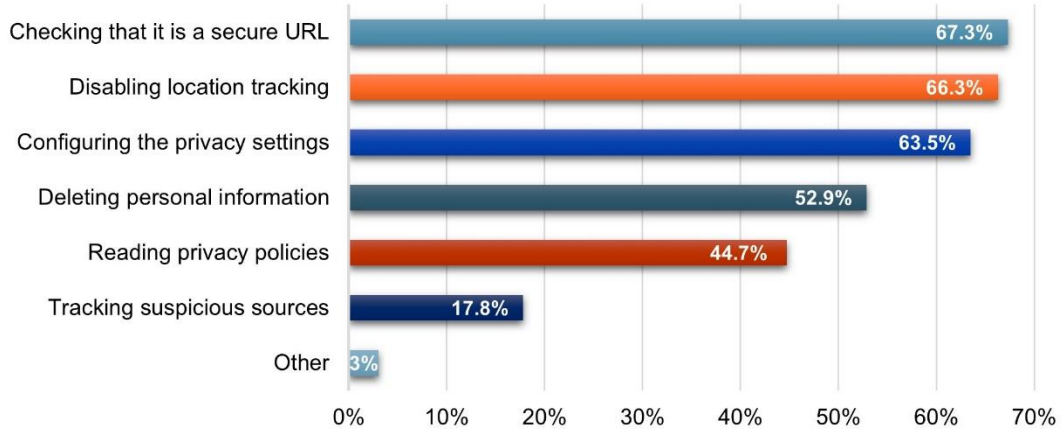
Source: Author's own work.

The Pearson chi-square statistic determines that there are significant differences in the results of items 6, 7, 8, and 9 concerning age. Users of older age ($X2 = 116.47$, $p<0.001$) are more concerned about their data being obtained by third parties (item 6) (Generation X, 93.6%; Millennials, 61%), as well as knowing how their data is used (item 9: $X2 = 83.59$, $p<0.001$) (Generation X, 95%; Millennials, 64%). On the other hand, Millennials express a higher level of knowledge about the treatment of their data (item 7: $X2 = 24.36$, $p<0.001$) and the use that organizations make of their information (item 8: $X2 = 23.41$, $p<0.001$).

Despite this evident concern, 40.4% of the sample indicates that they do not take actions to control and understand the treatment and use of their data in cyberspace (item 11). Those who do take measures report the following (item 12) (Figure 6): checking that it is a secure URL (67.3%), deactivating location tracking (66.3%), configuring profile privacy (63.5%), deleting personal information from the internet (52.9%), reading privacy policies (44.7%), and to a lesser extent, tracking suspicious sources before accessing them (17.8%). Additionally, 'Others' include actions such as blocking IP addresses, avoiding apps that demand excessive exposure, using access control systems on their computers, deleting unused services, and avoiding downloading unreliable apps. In this case, there are also significant differences in responses based on the age of the users, with Millennials ($X2 = 13.14$, $p<0.001$) being less cautious in protecting their data in cyberspace (50%) compared to subjects from Generation X (69.2%).

Figure 6

Responses to the item 'Actions of users to control the treatment and use of their data in cyberspace'.



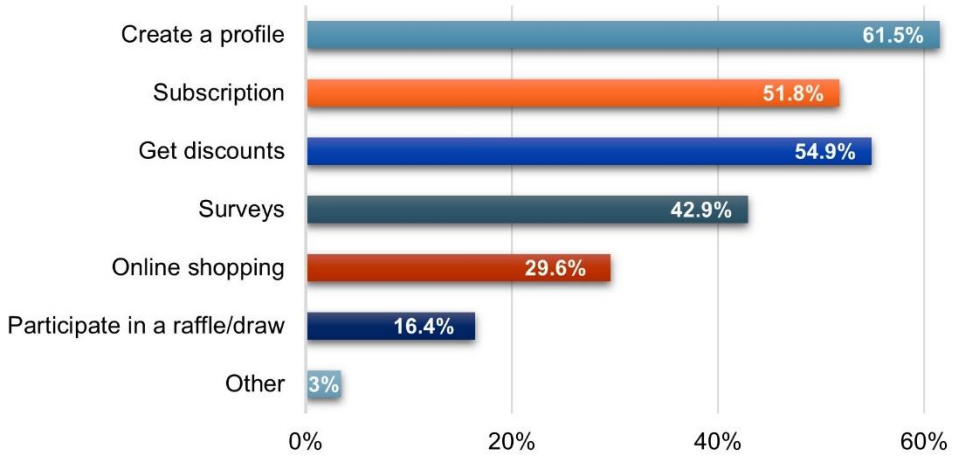
Source: Author's own work.

Considering the deliberate use of false information (item 13), more than half of the respondents (65.4%) claim to have provided them on some occasions, with varied intentions (item 14) (Figure 7): creating a profile (61.5%), subscribing to a service (51.8%), obtaining discounts or free samples (54.9%), receiving information about a product or service via email (54.9%), participating in a survey (42.9%), making online

purchases (29.6%), and entering a giveaway (16.4%). Additionally, there were 'Other' responses such as "responding to someone," "creating a profile to take online courses," or "registering on sites that I distrust".

Figure 7

Responses to the item 'Occasions when users have provided false information instead of their real data'.



Source: Author's own work.

Regarding the second dimension of the study (Table 2), Knowledge of Big Data, 83.7% of the users state that they are unaware of what it is and how it operates. Also, the majority (73.3%) believes that being a tool that uses data massively does not necessarily mean that it protects them. Only 20.3% of the sample is quite or completely in agreement with having to provide personal data to companies so that they can collect and use them for their own benefit in exchange for accessing certain services and applications. In fact, 76.8% are quite or very concerned that their data may be affected in some way due to a website security incident.

Table 2

Results of the study in the dimension of Knowledge of Big Data.

Response options	Total	Age		Gender	
		Millennials	Generation X	Man	Woman
I know how Big Data works.					
Not at all in agreement	59,3%	47,1%	71,5%	56,4%	61,9%
Slightly in agreement	24,4%	29,7%	19,2%	24,5%	24,3%
Fairly in agreement	12,5%	19,8%	5,2%	15,3%	9,9%
Completely in agreement	3,8%	3,5%	4,1%	3,7%	3,9%
Do you believe that Big Data, being a tool that uses data massively, protects your data?					
No	73,3%	55,8%	90,7%	71,2%	75,1%
Yes	26,7%	44,2%	9,3%	28,8%	24,9%
Are you agreeable to using services or applications in exchange for companies collecting and using your data?					
Not at all in agreement	50,6%	23,8%	77,3%	52,1%	49,2%

Slightly in agreement	27,3%	37,2%	17,4%	28,8%	26,0%
Fairly in agreement	13,1%	22,1%	4,1%	11,0%	14,9%
Completely in agreement	9,0%	16,9%	1,2%	8,0%	9,9%
Are you okay with big data collecting data about you to show you personalized advertisements in the future?					
Not at all in agreement	50,6%	30,8%	70,3%	51,5%	49,7%
Slightly in agreement	29,1%	36,0%	22,1%	27,0%	30,9%
Fairly in agreement	12,5%	18,6%	6,4%	13,5%	11,6%
Completely in agreement	7,8%	14,5%	1,2%	8,0%	7,7%
What do you think about your personal data being in the hands of others?					
It is a threat to my privacy.	61,0%	36,0%	86,0%	60,7%	61,3%
It is a mutual benefit for the user and the company.	28,2%	45,3%	11,0%	27,6%	28,7%
An opportunity to receive information related to my interests.	8,7%	15,7%	1,7%	8,6%	8,8%
Other	2,0%	2,9%	1,2%	3,1%	1,1%
Do you worry about your data being affected in a security incident when being processed by big data?					
Not at all in agreement	6,4%	8,7%	4,1%	5,5%	7,2%
Slightly in agreement	16,9%	25,6%	8,1%	18,4%	15,5%
Fairly in agreement	23,3%	40,1%	6,4%	20,2%	26,0%
Completely in agreement	53,5%	25,6%	81,4%	55,8%	51,4%

Source: Author's own work.

Regarding the fact that data is in the hands of other organizations, 61% of the surveyed users believe it is a threat to their privacy. Nearly 80% do not approve of continuously collecting information about their online activities to provide personalized promotional messages based on their specific interests. Despite this, 37% consider that big data can be seen as an opportunity in the sense that payments or internet searches result in more personalized contextual advertising. The open-ended response option has allowed adding the following contributions to the previous ones: "sometimes Instagram ads are interesting, but it's scary how much they control us," "I don't care because I mean nothing to these companies," "I don't see a problem with companies knowing which products I like," or "my proposal would be to omit personal data and be able to identify ourselves on the internet through a code."

Regarding the study of variable dependency, significant discrepancies have been detected according to age in almost all questions in this thematic block. Generation X users ($X2 = 27.11$, $p<0.001$) are the ones who are least familiar with the meaning of big data (item 15), as only 9.3% understand this concept compared to 23.3% of younger individuals or Millennials. Generation X subjects are also the ones who are least in agreement with the advantages and opportunities that big data offers in terms of protecting personal data (item 16: $X2 = 53.41$, $p<0.001$), allowing access to services (item 17: $X2 = 105.81$, $p<0.001$), or providing personalized content (item 18: $X2 = 62.18$, $p<0.001$); and they are the most concerned about incidents in internet security where their personal data may be involved or compromised (item 20: $X2 = 110.56$, $p<0.001$). This concern is expressed by 87.8% of them compared to 65.7% of Millennials.

5. DISCUSSION AND CONCLUSIONS

In the face of the increasing and more frequent collection of personal data by organizations, online privacy has become an urgent concern for citizens, requiring

investigation by academic institutions (Schomakers et al., 2019). This has been precisely the intention of our work, and the results show that the main objective of exploring users' opinions and behavior regarding the treatment and use of their data by third parties, as well as the advantages and risks associated with these processes and their relationship with the phenomenon of big data, has been fulfilled.

The majority of citizens, almost 84%, are unaware of what data intelligence is and how it works. In general, they feel unprotected and do not agree with having to provide so much information to access internet services and applications, considering it a threat to their privacy. However, nearly 40% view big data as an advantage in terms of receiving more personalized content according to their interests, although in practice, this personalization is not clearly perceived, as determined by the research of Martín-Herrera and Guerrero-Solana (2019).

Despite the concerns identified, a significant portion of the sample, especially younger individuals, state that they do not take specific actions to control access to their data by third parties. These results, which align with the findings of Tabassum et al. (2019), support the first hypothesis of the study, affirming that there is a prevailing lack of knowledge, and users do not approach the sharing of their data in cyberspace with enough caution. In fact, a study by Sánchez-Holgado et al. (2022) observes that a high percentage of Spanish citizens believe they receive little or very little information about data science, even though the trust they declare in this information is high.

Another objective of the research was to determine how gender and age influence users' perceptions regarding the study subject. In this regard, the analysis of dependence using the non-parametric Pearson's chi-square test allows us to reject the second hypothesis and confirm the third one, as significant differences have only been observed in the results concerning the personal variable of age, which shows a statistically significant relationship with most of the questionnaire's questions. On the other hand, gender does not cause any discrepancies in any of the proposed questions. Illustratively, it is worth noting that older subjects, belonging to the so-called Generation X, are more hesitant and concerned about processes or procedures involving the sharing of personal data on the internet. In contrast, younger individuals or Millennials, aged between 19 and 39 years, demonstrate greater knowledge about the functioning of big data, perceiving fewer risks in managing this information, and recognizing more advantages, such as access to personalized services and content.

Regarding this generation gap, which has been detected in studies analyzing the attitudes and behaviors of different user generations on the internet, there is a certain technological determinism prevailing in a discourse that defends the idea that young people who have been born into a digital world have radically different behaviors (Matellanes-Lazo, 2012, p. 18). However, beyond the skills in handling devices and electronic tools, other aspects such as the education received or current social imperatives must also be taken into account to explain these differences.

Regarding the implications of the study, this diagnosis can be taken into account for designing educational campaigns targeted at users in general or specific groups, such

as educators (Correa-Gorospe et al., 2021). Through digital literacy programs, users can learn to manage their data correctly and securely, and become aware of the risks associated with poor practices in an "era of growing datafication, platforms, algorithmic decision-making, and digital transformation" (Reilly et al., 2022, p. 85), where "numerical evaluations and the volume of comments are a form of informational social influence" (Gavilán-Bouzas et al., 2018, p. 89).

Similarly, the results can be useful for legislators and web developers, enabling them to address the identified issues from their respective fields, as data protection on the internet remains one of the challenges of the digital society. Ideally, data mining techniques for preserving privacy should extract data without compromising the security of people's confidential information, particularly at the level of record (Binjubeir et al., 2019). Additionally, "minimum standards of information and freedom necessary for individuals to control the use of data collected through tracking technologies" should be met (González-Guerrero, 2019, p. 209).

Finally, it should be noted that the development of this work has inspired new research directions. These may include repeating the study to understand how attitudes of citizens change over time and with technological advancements, working with a larger sample, as obtaining a greater number of questionnaires during the fieldwork would have been desirable, or comparing the results with those from other countries to identify common patterns and understand how cultural impact influences users' perceptions and behaviors.

6. REFERENCES

Agencia Española de Protección de Datos. (2018). *Protección de datos y prevención de delitos*. https://acortar.link/y9JkKC

Binjubeir, M., Ahmed, A. A., Ismail, M. A. B., Sadiq, A. S. y Khan, M. K. (2019). Comprehensive survey on big data privacy protection. *IEEE Access*, 8.

Cerezo-Domínguez, A. I. y García-Cornejo, R. (2020). La ciberdelincuencia en España: un estudio basado en las estadísticas policiales. *Revista Peruana de Ciencias Penales*, *1*(34), 91-106. https://doi.org/10.56176/rpcp.34.2022.3

Comisión Europea. (2020). *Estrategia europea de datos*. https://acortar.link/rdQCfp

Correa-Gorospe, J. M., Losada-Iglesias, D. y Gutiérrez-Cabello Barragán, A. (2021). Big Data y la alfabetización posthumana del futuro profesorado. *Sociología y tecnociencia: Revista digital de sociología del sistema tecnocientífico*, *11*(2), 102-122.

Escobar, M. y Mercado, M. (2019). Big data: un análisis documental de su uso y aplicación en el contexto de la era digital. *Revista La Propiedad Inmaterial*, 28, 273-293. https://doi.org/10.18601/16571959.n28.10

Fundeu. (2013). *Macrodatos e inteligencia de datos, alternativas a big data*. https://www.fundeu.es/recomendacion/macrodatosalternativa-abig-data-1582/

Gavilán Bouzas, D., Martínez Navarro, G. y Fernández Lores, S. (2018). Comunicación en el sector turístico. Puntuaciones y comentarios de usuarios como variables estratégicas. *Vivat Academia. Revista de Comunicación*, 144, 77-94. http://doi.org/10.15178/va.2018.144.77-94

González, A. (2015). *Privacidad en Internet: Los derechos fundamentales de privacidad e intimidad en Internet y su regulación jurídica. La vigilancia masiva* [Tesis Doctoral, Universidad de Castilla-La Mancha]. https://bit.ly/3MqdkJE

González-Guerrero, L. D. (2019). Control de nuestros datos personales en la era del big data: el caso del rastreo web de terceros. *Estudios Socio-Jurídicos*, *21*(1), 209-244.

Hassan-Montero, Y. (2006). Factores del Diseño Web Orientado a la Satisfacción y No-Frustración de Uso. *Revista Española de Documentación Científica*, *29*(2), 239-257. http://redc.revistas.csic.es/index.php/redc/article/viewFile/291/353

Hernández-Leal, E. J., Duque-Méndez, N. D. y Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, *20*(39). https://doi.org/10.22430/22565337.685

Hernández Sampieri, R., Fernández-Collado, C. y Baptista, P. (2006). *Metodología de la Investigación*. McGraw Hill.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group*, 949. https://bit.ly/3DTvnE4

LOPDGDD. (2018). *Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales*. https://www.boe.es/eli/es/lo/2018/12/05/3

Martín-Herrera, I. y Guerrero-Solana, L. (2019). La influencia del Big Data en el email marketing. Un estudio desde la percepción de los usuarios. En A. Rey, C. Martínez, G. Albort y A. L., Leal (Coord.), *Libro de Actas del II Congreso Iberoamericano AJICEDE* (pp. 356-362).

Matellanes-Lazo, M. (2012). Actitudes, comportamientos y usos de diferentes generaciones de usuarios en Internet. *Vivat Academia. Revista de Comunicación*, 115, 1-22. https://doi.org/10.15178/va.2011.115.1-22

Meraz-Espinoza, A. I. (2018). Empresa y privacidad: el cuidado de la información y los datos personales en medios digitales. *Revista IUS*, *12*(41), 294-308. https://doi.org/10.35487/rius.v12i41.2018.313

Parlamento Europeo. (2017). *Resolución del Parlamento Europeo sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225 - INI).*

Parlamento Europeo. (2018). *Privacidad de la protección de datos*. https://www.europarl.europa.eu/privacy-policy/es/data-protection

Real Academia Española. (s.f.). *Privacidad*. En Diccionario de la Lengua Española. https://dle.rae.es/privacidad?m=form

Reilly, K., Flores, M. y Morales, E. (2022). Aprendiendo sobre Alfabetización de Datos Personales: criterios para evaluar intervenciones en América Latina. *Observatorio (OBS*)*, 16, 85-103.

Rubio-Laborda, J. F., Almansa-Martínez, P. y Pastor-Bravo, M. M. (2021). Relaciones sexistas en la generación X y Millennials. *Atención Primaria*, *53*(4), 1-10, https://doi.org/10.1016/j.aprim.2021.101992

Sánchez-Holgado, P., Arcila Calderón, C. y Blanco-Herrero, D. (2022). Conocimiento y percepción de la ciudadanía española sobre el big data y la inteligencia artificial. *ICONO 14. Revista científica de Comunicación y Tecnologías Emergentes*, *20*(1). https://doi.org/10.7195/ri14.v20i1.1908

Schomakers, E. M., Lidynia, C., Müllmann, D. y Ziefle, M. (2019). Internet users' perceptions of information sensitivity–insights from Germany. *International Journal of Information Management*, 46, 142-150.

Suras, S. (2020). *Principales riesgos digitales y cómo prevenirlos*. https://bit.ly/3rJGBHp

Soto, Y. (2017). Datos masivos con privacidad y no contra privacidad. RBD. *Revista de Bioética y Derecho*, 40, 102-113. https://doi.org/10.1344/rbd2017.40.19165

Tabassum, M., Kosinski, T. y Lipford, H. R. (2019). "I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)* (pp. 435-450).

Zúñiga-Becerra, O. (2018). Educación y prevención en materia de protección de datos personales de niños, niñas y adolescentes en internet. *Estudios en derecho a la información*, 5, 59-79.

7. Related articles

Aladro Vico, E. (2020). Comunicación sostenible y sociedad 2.0: particularidades en una relación de tres décadas. *Revista de Comunicación de la SEECI*, 53, 37-51. https://doi.org/10.15198/seeci.2020.53.37-51

Brandariz Portela, T., Ortega Fernández, E. y Padilla Castillo, G. (2023). Los millenials y la honestidad como valor de marca. El caso Goiko en Instagram. *Revista de Comunicación de la SEECI*, 56, 214-238. https://doi.org/10.15198/seeci.2023.56.e820

Caldevilla-Domínguez, D., Barrientos-Báez, A. y Parra-López, Eduardo (2020). Horizontes del mundo digital: de la simulación y la banalización de la experiencia, un uso social, ecológico e innovador de la Sociedad Red. *Cuadernos de Información y Comunicación,* 25, 269-277. http://dx.doi.org/10.5209/ciyc.68722

Carrasco Polaino, R. y Villar Cirujano, E. (2021). Greta Thunberg como personaje viralizador de los tuit del sector informativo durante la cumbre del clima COP25. *Revista de Ciencias de la Comunicación e Información*, 26, 1-13. https://doi.org/10.35742/rcci.2021.26.e116

Sáenz Barrera, A. G., Ganga-Contreras, F. A. y Marañón Lazcano, F. M. (2022). Representación del Millennial en Latinoamérica. Elementos preliminares que definen a una generación. *Revista Opción*, 97, 154-171.

AUTHORS' CONTRIBUTIONS, FUNDING, AND ACKNOWLEDGEMENTS

Authors Contributions:

Conceptualization: Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes. **Methodology:** Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes. **Software**: Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes. **Validation**: Guerrero Caballero, Mercedes. **Formal Analysis**: Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes**. Data curation**: Guerrero Caballero, Mercedes. **Writing-Preparation of the original draft:** Guerrero Caballero, Mercedes. **Writing-Revision and Editing**: Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes **Visualization** Guerrero Caballero, Mercedes. **Supervision:** Martín Herrera, Inmaculada y Guerrero Caballero, Mercedes. **All authors have read and accepted the published version of the manuscript.**

AUTHOR(S)

Inmaculada Martín Herrera: Doctor in Education Sciences. Master's in Communication. Bachelor's in Journalism. Research professor in Communication and Digital Communication degrees at the San Isidoro University Center, affiliated with the Pablo de Olavide University in Seville, Spain. Director of the magazine La Crónica de Sevilla. Research areas: online journalism, educational technology, and media and digital literacy.
Orcid ID: http://orcid.org/0000-0002-8943-4766
Google Scholar:
https://scholar.google.es/citations?view_op=list_works&hl=es&hl=es&user=puchrOYAAAAJ
ResearchGate: https://www.researchgate.net/profile/Inmaculada-Martin-Herrera
Academia.edu: https://independent.academia.edu/InmaculadaMart%C3%ADnHerrera

Mercedes Guerrero Caballero: Graduated in Digital Communication from the San Isidoro University Center, affiliated with the Pablo de Olavide University in Seville, Spain. Master's in Professional SEO and SEM from Kschool (School of Internet Professionals) in Madrid. SEO and SEM professional and social media manager in national-level organizations. Her research areas include digital marketing, web positioning, big data, online privacy, and the Internet of Things.
Orcid ID: https://orcid.org/0009-0001-7438-5498
Google Scholar: Scholar:
https://scholar.google.es/citations?hl=es&user=b8vgn_gAAAAJ&view_op=list_works&gmla=ABEO0YowbAymPXB16WsxSoYCRIKcQWlzggGITTuyoqso7IgcmMiHZD8mxG1bQtzMe2ddQBEuOukFnppSSHVqKY62QsM_JNj3ADlwfpbi