Data Science in the Age of Artificial Intelligence: Challenges and Solutions

Avin javadi shejooni[1]

Islamic Azad University of Rasht

Abstract

In the era of rapid advancements in Artificial Intelligence (AI), both corporate and public sectors have experienced transformative shifts in their organizational models. As AI becomes increasingly integrated into economic systems, regulatory frameworks, and decision-making processes, it underscores the critical role of data science in navigating the complexities of a globally connected society. This paper examines the emerging concept of Behavioral Data Sciences (BDS) as a multidisciplinary approach to understanding and predicting human behavior through AI. While BDS offers significant benefits, such as improved accuracy in predictions and insights into social dynamics, it also raises profound ethical concerns, particularly regarding user privacy and the potential for surveillance capitalism. Through an exploration of current governmental AI strategies and case studies, this study seeks to address the ethical challenges and propose solutions for the responsible use of AI in public governance.

Keywords: Data sciences, Governments, Artificial intelligence, Surveillance capitalism

1. Introduction

In recent years, the rapid development of Artificial Intelligence (AI) has spurred significant transformations in organizational models across both corporate and public sectors (Brynjolfsson & Mitchell, 2017). Within a globally connected society where the Internet serves as the primary communication tool, the integration of AI into economic systems, regulatory frameworks, and decision-making processes has become increasingly prevalent. These developments have underscored the importance of data and behavioral analysis as critical elements for public actors to navigate complex social dynamics (Ballestar, Camiña, Díaz-Chao, & Torrent-Sellens, 2021; Irvin & Stansbury, 2004).

In this context, the conceptualization and establishment of theoretical and legal frameworks that delineate ethical boundaries for the analysis, treatment, and utilization of citizens' data pose significant challenges. These challenges span scientific, legal, and professional domains, as the implications of AI-driven insights into human behavior continue to raise concerns regarding user privacy and autonomy (Kamolov & Teteryatnikov, 2021; Narayanan, Huey, & Felten, 2016). The concept of Behavioral Data Sciences (BDS) has emerged as a critical interdisciplinary field aimed at addressing these issues by combining methodologies from behavioral sciences with advanced data science techniques to understand and predict human behavior using AI (Saura, Palacios-Marqués, & Iturricha-Fernández, 2021).

This paper explores the implications of BDS within governmental strategies, highlighting the potential risks to privacy and individual freedoms posed by AI-driven behavioral predictions. The rise of surveillance capitalism, where user-generated data is commodified and leveraged for economic gain, further complicates the ethical landscape surrounding AI and BDS (Zuboff, 2019b). Through an examination of case studies and current practices, this study seeks to shed light on the ethical considerations that must guide the integration of AI and BDS in public governance.In parallel to the increase in the use of AI in governments, and as a consequence of the evolution in data and behavioral analysis practices, the concept of behavioral data sciences (BDS) has been developed to combine a multitude of issues related to data science and behavior (Harari et al., 2016). Although the very term “behavioral data sciences” does not appear in the scientific literature, several previous studies, including Agarwal and Dhar (2014) and Van Der Aalst (2016), have directly defined the future guidelines for its development. Therefore, the term BDS refers to a new and emerging interdisciplinary field that combines techniques from behavioral sciences, psychology, sociology, economics, and business, and uses the processes from computer science, data-centric engineering, statistical models, information science, or mathematics, in order to understand and predict human behavior using AI (Saura, Palacios-Marqués, & Iturricha-Fernández, 2021). In essence, BDS is a mix of disciplines that combines knowledge of the data that users or citizens publicly generate on the Internet -known as user-generated content (UGC) or Data (UGD) through the use of mobile applications and other connected devices, such as Internet of Things (IoT), smart homes, self-driven cars, or through smart-cities connected services (Schreiner, Fischer, & Riedl, 2019).

With the use of techniques focused on BDS, governments could apply algorithms that work with AI and systems that analyze behavior (Grimmelikhuijsen, Jilke, Olsen, & Tummers, 2017), identify patterns to explore the knowledge about the society (Men & Tsai, 2014) as well as its consumers or users (Chen et al., 2020; Chen et al., 2021). In this study, the use of the BDS concept is specifically linked to the analysis of AI strategies developed by governments to date. Since the term does not relevantly appear in the published scientific literature, the present study is pioneering and original in this respect.

Furthermore, corporates not only leverage users and clients' data to improve their products and services, but also use them as exchange currency with interested third parties, such as governments or other public institutions (Silverman, 2017). Therefore, by studying user behavior data, companies and governments develop sophisticated power machines that predict an economic logic that helps corporates generate more money at the expense of users and citizens (Zuboff, 2019a). Likewise, according to the government actions, the use of AI raises concerns about privacy and personal security issues (Yang, Elisa, & Eliot, 2019). While predictions are not equal to observations, the more data is obtained from the society, the greater is the ability to predict. Accordingly, predictions can reach the same level of effectiveness as that of observations (Zuboff, 2019a). Therefore, if governments have this intelligence, and if it is also automated based on AI, the risk to privacy and free decision-making in the society could be at threat (Mazurek & Małagocka, 2019).

In this context, a key notion in this field is the concept of surveillance capitalism. According to Cinnamon (2017) and Zuboff (2019b), in surveillance capitalism, user experience and behavioral data are used as economic drivers to create a new economy where economic drivers and profits come from predicting how users behave. Therefore, considering this new concept, governments can take action and use AI as a tool focused on BDS. However, as stated Bromberg, Charbonneau, and Smith (2020), such use can violate citizens' privacy and security. For example, by using AI and BDS, governments can interfere with the behavior of the society to achieve a change in behavior, without the society being aware of it (Zuboff, 2015). There is also evidence of how governments can use AI to predict election results using massive data to change the voting intentions of thousands of users (Isaak & Hanna, 2018). This was the case of US Facebook users' behavioral data that, when analyzed with behavioral prediction algorithms, such as the one developed by Cambridge Analytica (Heawood, 2018), were employed to modify the election results in the US presidential campaign between Donald Trump and Hilary Clinton in 2016 (Cadwalladr & Graham-Harrison, 2018). Another example could be the famous German doll, Cayla, which recorded chunk dialogues said by children (Haynes, Ramirez, Hayajneh, & Bhuiyan, 2017), the company then sold those data Nuance Communications, which, in turn, developed a voice recognition software and sold it to the US Central Intelligence Agency (CIA) (Madnick, Johnson, & Huang, 2019). In this case, we can speak about government suppliers' provision of AI-related services that are unethical from citizens' point of view.

In this context, after the development of this type of events where AI, governments, and the data collection capacity of corporations is questioned, essential questions regarding the knowledge, authority, and power of government use of BDS techniques should be explored. Furthermore, understanding the predictive ability that government institutions might obtain if they train AI models that can predict user behavior is a prerequisite for any society to feel confident about implementing new technologies (Hobolt, Tilley, & Wittrock, 2013). Of note, data predictions and models that work with the prediction of human behavior are becoming dominant forms of

capitalism and generate new business models and new products in the form of data (Zuboff, 2019b). Of note, BDS is a clear priority for the development of ethical strategies by governments when they implement AI in their strategies as it is presented as a new concept linked to user privacy, AI deployment in governments, or behavioral analytics, that brings together all of the above in the form of analysis of society's behavioral data.

However, several unanswered questions remain, such as what is the legitimacy of predicting user behavior? And who do these behavioral data belong to? Based on the privacy concerns outlined above and the originality of the study justified under the BDS new emerging concept, to the best of our knowledge, none of previous studies had identified and described the risks of governmental implementation of AI to citizens' privacy. Furthermore, there has been no research linking the concept of BDS to the main uses of AI by governments. Thus, we seek to fill a gap in the literature by exploring the possible uses and risks to citizens' privacy if governments implementation of AI in their strategies under the new BDS conceptual framework. To this end, this study first develops a systematic review of the literature to establish and confirm the main scientific contributions to date in this field of study. Secondly, based on the results of the systematic review of the literature, 15 interviews were conducted with 11 individuals working in the government; of these, 2 were economists for the government, and 2 belonged to organizations that advise the government. Thirdly, based on the coded results of the interviews, two data-mining techniques (topic-modeling and textual analysis) were developed to identify insights and create knowledge related to the object of study. Following this approach, the present study aims to identify and discuss the main practical and theoretical implications for governments when using AI-based strategies with BDS techniques.

Therefore, in order to cover the identified gap in the literature, the present study addresses the following research questions (RQ): RQ1: What kind of citizens' privacy issues are expected when governments use behavioral-based AI in their strategies? and RQ2: What AI techniques can governments develop to predict the society's behavior?

Based on the results, we discuss theoretical implications regarding the application of AI strategies used by governments that respect the privacy of citizens' data. In addition, the main contributions to date are theorized in relation to the management of user data and the need to regulate security, ethics, and privacy of user data. Similarly, we also discuss practical implications that form a guide for the application of AI strategies by governments that avoid any type of privacy violations linked to surveillance capitalism actions.

The remainder of this paper is structured as follows. In Section 2, the theoretical framework of the study is presented. Section 3 discusses the methodological approaches used. Section 4 reports the results. Section 5 provides a discussion of important theoretical contributions and future directions that our results offer for the analysis of BDS privacy issues in government AI deployment. Conclusions, along with a discussion of theoretical and practical implications.

2. Theoretical framework

2.1. Understanding surveillance capitalism and behavioral data sciences

As argued by Zuboff (2019b) and Belhadi et al. (2021), we are living in one of the deepest transitions in the information age—namely, in an ecosystem where data are the largest source of

information. Seeking to outline a theoretical background with the main concepts used to analyze and predict user behavior in the digital ecosystem, this section identifies the main theoretical perspectives used in the literature to analyze the factors that contribute to the development of AI in governments.

For their part, governments need to be updated and use the latest technologies to understand what the demands of the society are (Figenschou, 2020). However, according to many initiatives, the regulation of the Internet itself is not working, and the society demands that its data should remain anonymous at all costs (Zuboff, 2019b). This raises concerns about user privacy (Ribeiro-Navarrete, Saura, & Palacios-Marqués, 2021). Users are aware of the fact that, based on the analysis of human experiences linked to behavioral data, governments can turn their actions into sophisticated intelligent machines capable of predicting any issue targeted by governments (Kavanaugh et al., 2012). Therefore, the ultimate goal will always be to understand the future behavior of the society regulated by governments (Linders, 2012).

Under this paradigm of privacy concerns about AI and its implementation by governments to monitor, actively listen, trace possible states of alarm, or predict any kind of event that negatively affects the society, the concept of surveillance capitalism is born (Cinnamon, 2017; Zuboff, 2015; Zuboff, 2019b). The concept of surveillance capitalism advocates that human experience is unilaterally automated as data sources to predict human behavior (Zuboff, 2019a). While there are indeed objectives of service improvement and understanding the society's behavior to improve the public offer by governments (Andrew & Baker, 2019), the concept of surveillance capitalism also implies that humans are used as products of massive data production to improve the economic profitability of companies at the expense of the data about user behavior (Zuboff, 2015).

In the circumstances where ethical actions are lacking, companies and governments use behavioral data to make the society behave in ways that are more convenient to obtain greater economic benefits (Zuboff, 2015). When viewed from a business perspective, this leads to an increasing number of Internet-centered business models that cater to addictive behavioral patterns (Hou, Xiong, Jiang, Song, & Wang, 2019). In this way, users generate more data about their behavior; accordingly, their attitudes and feelings can be predicted. Then, based on these actions, companies and governments generate more profitability on the advertising (Palos-Sanchez, Saura, & Martin-Velicia, 2019) products shared in these business models (Dwivedi, Kapoor, & Chen, 2015), or using user behavior data as the basis of data-centered strategies (Dwivedi et al., 2018).

In surveillance capitalism, the main source of data is the information generated by users while using connected devices. All this information is analyzed using BDS that takes a new perspective of analysis through a combination of different fields of research (Zhuoxuan, Yan, & Xiaoming, 2015). In recent years, the number of tools used by both governments and companies to obtain data has considerably increased (Paul & Aithal, 2020). In fact, many variables indicate parameters for measuring user behavior on the Internet or through their mobile and connected devices (Hobolt et al., 2013).

Until now, the main sources of data were websites, cell phones, intelligent organization systems, Customer Relationship Management (CRM) systems, and marketing automation sources, among others. This type of data always generates categories known as events or objectives, which have the purpose of explaining some properties defined by the organizational structure of the data-analysis system (Abou Elassad, Mousannif, Al Moatassime and Karkouch, 2020). However, as mentioned above, the number and type of connected devices has recently exponentially increased,

from IoT to smart city services, among other connected devices (Kankanhalli, Charalabidis, & Mellouli, 2019).

The understanding of user behavior data on the Internet has led to the emergence of new digital marketing strategies in the business ecosystem (Dwivedi et al., 2020). It is not the first time that the business ecosystem offers opportunities and benefits to government institutions to maximize their processes (Zhang, Wang, & Zhu, 2020), increase the efficiency of their strategies (Pencheva, Esteve, & Mikhaylov, 2020), or create new listening tactics (Macnamara, 2015). Following these considerations, Table 1 presents the main concepts related to BDS analysis that can be used by governments to monitor user behavior through the data they generate.

Table 1. Main concepts linked to the analysis of behavioral data.

Concept	**Description**	**Authors**
Behavioral Data Sciences (BDS)	An interdisciplinary field that studies user behavior through the data they generate from sociological, psychological, and economic perspectives applying statistics, mathematics, and data automation.	Litman, Robinson, and Abberbock (2017) Xu, de Barbaro, Abney, and Cox (2020)
Machine Behavior (MB)	A field that leverages behavioral sciences to understand the behavior of AI agents.	Abou Elassad, Mousannif, Al Moatassime and Karkouch (2020) Oey, Jones, Bullard, and Sant (2020)
Algorithmic Behavior (AB)	A field of study of the BDS using algorithms in large databases	Hobolt et al. (2013) Macnamara (2015)
Behavioral Analytics (BA)	Study of user behavior data using the Internet and social networks	Touma, Bertino, Rivera, Verma, and Calo (2017) Khan (2017)
Behavioral Economics (BE)	Studies the effect of the cognitive and emotional psychology of culture and society on the predictions of economic theory	Hursh (1984) Streletskaya et al. (2020)
Big Behavioral Data Science (BBDS)	Refers to BDS and the use of Big Data techniques	Gomez-Marin, Paton, Kampff, Costa, and Mainen (2014)
Behavior Informatics (BI)	Investigates user behavior data with processes focused on computer sciences.	Cao et al. (2014) Paul and Aithal (2020)

Concept	Description	Authors
Collective Behavior Analysis (CBA)	Various approaches to the study of behavior focused on the collective knowledge of individual behavior	Belhadi et al. (2021) Turner and Killian (1957)
Behavior Learning Analysis (BLA)	A scientific field for the study of behavior with data entry learning techniques	Zentall, Galizio, and Critchfield (2002) Zhuoxuan et al. (2015)

2.2. Main user's behavior data sources used by the government in their monitoring strategies

In the ecosystem that drives the development of the economy based on behavioral analysis and its data, the importance of data sources can hardly be overestimated (White & Boatwright, 2020). As argued in many previous studies, since users are not fully aware that their data will be used and sold to interested third parties for a financial contribution (Acar, Englehardt, & Narayanan, 2020), ethics is not an essential component of new business models focused on massive data collection and analysis (Löfgren & Webster, 2020).

Moreover, several available initiatives such as the new GDPR legislation introduced by the European Union through the European Commission to protect users based on abusive uses of their data are insufficient (Sørensen & Kosta, 2019). Although the new regulation obliges companies to explicitly specify how the data are used, in reality, users do not possess knowledge necessary to understand privacy policies and legal notices of the applications on their mobile computers and any types of connected device (Martin, 2015).

There is evidence that, due to the psychological phenomenon known as “instantaneous reward”, despite some awareness about privacy issues, the millennial generation and the digital natives prefer to use the applications as fast as possible instead of taking time to understand how their data will be used (Hull et al., 2004). In many situations, including the recent Covid-19 pandemic, the issue of user privacy has raised many concerns (Maher, Hoang, & Hindery, 2020). During the Covid-19 crisis, in order to track Covid-19 infections and notify citizens if they have been in contact with an infected, many governments have decided to ask citizens to use applications that track their location (Gerard, Imbert, & Orkin, 2020).

A recent analysis of these novel monitoring techniques by governments suggested that citizens frequently use this type of active listening (Maher et al., 2020; Zhou, Yang, Xiao, & Chen, 2020). The sources of data that governments may have access have been studied previously within several projects, such as the one published by The New York Times (Thompson & Warzel, 2019). Specifically, Thompson and Warzel (2019) highlighted many decision-making concerns that governments may have with access to multiple companies collecting information from users. Then, governments use user data to improve their processes of monitoring the society and its behavior (Thompson & Warzel, 2019), thereby prioritizing the issue of national security. With this type of strategy, user behavioral data are used as a source that governments use to train their algorithms that work with machine learning. Accordingly, behavioral data analysis is of a paramount important for governmental strategies focused on AI (Ribeiro-Navarrete et al., 2021).

Of note, user data can be transferred by third parties to governments (Thompson & Warzel, 2019).

Systematic review of the literature

To better understand the main uses of AI by governments as studied in the scientific literature to date, we conducted a systematic review of the literature (de Camargo Fiorini, Seles, Jabbour, Mariano, & de Sousa Jabbour, 2018). Systematic literature reviews are exploratory research approaches used to understand emerging new fields of study (Kraus, Breier, & Dasí-Rodríguez, 2020). A major reason underlying the recent increase in the number of systematic literature reviews is that a literature review makes it possible to outline a theoretical framework with the main agents that contribute to the development of the proposed research objective. Therefore, the aim of systematic literature review is to analyze an emerging issue and to identify the main techniques employed to study that issue. Therefore, systematic reviews are an effective method to identify the proposed objectives related to AI uses in governments and citizens privacy (Zuiderwijk et al., 2021).

In the present study, we followed the procedure developed by Bem (1995), who proposed that a systematic review should be divided into the following three steps. In the first step, the topics to be discussed within the scientific area are identified. To this end, keywords are identified that can summarize the objective of the research through searching databases (Sarkis, Zhu, & Lai, 2011). In the second step, the searches in these databases are performed, the collected data are filtered, and the results are analyzed (Akter & Wamba, 2016). During the filtering process, titles, abstracts, and keywords of potentially relevant studies are examined. This is followed by the analysis of the content of the articles, and their suitability for the review is assessed. The studies that do not meet these criteria are excluded from the systematic review process. In the third step, the content of the contributions retained in the sample is analyzed, and the main concepts are discussed (Zeng, Hu, Balezentis, & Streimikiene, 2020).

In the present study, final contributions were selected during the review process that focused on identification of the main purposes of each potentially relevant study (Akter et al., 2019). The searches were conducted in the following databases: Web of Sciences (WOS), IEEE Xplore, ScienceDirect, ACM Digital Library, and AIS Electronic Library. The keywords used to search the databases were “Government” OR “Governance” OR “Public Management” OR “Public Sector” OR “Public Administration” OR “Public Policy” OR “State” OR “Municipality” OR “Citizens” AND “Artificial Intelligence” OR “AI” OR “Predictive Analytics” OR “Intelligence Systems” OR “Expert Systems” OR “Collective Behavior” OR “Surveillance Capitalism” OR “Behavioral Analysis”. The searches were performed between October 5 and 10, 2020 and updated in January 2022. Of note, the search term BDS has not been used in this process, since the results of government studies using AI were analyzed from the perspective of BDS as a new emerging concept, which this study thoroughly outlines and defines in the results.

The results of the process were as follows. In WOS, 20 articles were selected from a total of 65 potentially relevant results; in ScienceDirect 7 results were selected from a total of 29 studies; in AIS Library, the total number of potentially relevant studies was 3, of which only 1 was retained in the final dataset; ACM Digital Library, a total of 4 studies were found, of which 2 were selected; finally, in IEEE Xplore, of a total of 20 potentially relevant study, 4 were selected. Therefore, after the selection process, a total of 34 research studies were selected to be included in the present study. For the exclusion criteria, we followed PRISMA evidence-based minimum set of items (Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021a) aimed to filter quality research studies. First, the abstracts and keywords of the articles were analyzed to identify inadequate and not inclusive

terms related to the objectives of the study. Second, an in-depth analysis of the articles identified as suitable was performed. Next, we analyzed whether the objectives of the study were directly or indirectly linked to the objectives of the present research.
ed below.

In-depth interviews

Seeking to obtain additional knowledge regarding the uses of AI by governments and the concerns related to citizens' privacy, we conducted in-depth interviews with informants working in governments. Following the guidelines proposed by MacDougall and Fudge (2001), our qualitative interviews were held with politicians, senators, and other government-related officials in Spain.
The ultimate goals of these interviews was not to quantitatively assess the studied phenomenon, but rather to gain a deep understanding of it by obtaining information from an original primary source. The importance of such qualitative approach was previously justified by Orlikowski and Baroudi (1991) and Roberts (2015). Subsequently, the content of the interviews was used to build theory and extract insights.
We conducted a total of 15 interviews on user privacy and AI strategies developed by governments. Of these 15 interviews, 5 were conducted by phone (Pell et al., 2020), 2 by video call (Lukacik, Bourdage, & Roulin, 2020), 4 in person (Lukacik et al., 2020), and 4 by email (McKinley, Fong, Udelsman, & Rickert, 2020). In all cases, the interviews were digitally coded for further analysis under the Natural Language Processing (NLP) framework. Of 15 informants, 11 individuals worked in the government, 2 were economists for the government, and 2 belonged to organizations that advise the government (see Table 7). The informants were Spanish (12), Venezuelan (1), Egyptian (1), and Colombian (1) nationals. Their identities are anonymized in the present study (Natow, 2020). The three interviewees who were not native Spaniards live in Spain and work for governments or corporations linked to economics, finance, and politics. All members taking part in the interviews are linked to the Club Financiero Génova (CFG) in Madrid, a club focused on economic development, business, and politics. The interviewees were informed of the interviews at various events held at the CFG and contacted afterwards. The interviews were conducted in Spanish and translated into English.
Of note, as informed by the European Commission, Spain has developed a strategy report to monitor the development, as well as to uptake and measure the impact of AI in their government actions. The Spanish government has informed that they use AI to facilitate the development and deployment of the economy and society. Its strategy adopts a multidisciplinary approach to address economic, social, environmental, public management, and governance challenges, and it includes perspectives for a wide range of sectors and disciplines (European Commission, 2020).
In-person interviews and video call interviews lasted for about 30–40 min each. Telephone interviews averaged 20–25 min in length. Email interview responses averaged 750–600 words each. Interview data were collected between October 15, 2020, and January 8, 2021. Questions are shown in Appendix A. The informants were selected based on the work they do or have previously done in the government. All informants were linked to public administrations, governments, political parties, or advisors to the government. Our interviews were semi-structured and included open-ended questions. Table 4 shows the characteristics of our informants based on their role, industry of specialization, professional status, organization they belong to, and nationality.

Table 4. Interviewees by role of informant, industry, professional state, organization, and

In the last decade, data-mining techniques have been extensively used notably in the scientific literature (Yang & Wu, 2006). These techniques are used to create knowledge and extract insights from both structured and unstructured databases (Wu et al., 2003). A combination of several data-mining techniques processes can provide truly relevant insights into the objects proposed under study (Jindal & Borah, 2013).

In the present study, two data-mining processes were combined: Latent Dirichlet Allocation (LDA) and Textual Analysis (TA). The first one was a topic-modeling algorithm developed in Python to extract insights in the form of topics. LDA was applied to the database containing the content of the in-depth interviews (Blei, Ng, Jordan, & Lafferty, 2003; Pritchard, Stephens, & Donnelly, 2000). The novelty of this approach is that we used a methodology typically applied to analyze to explore primary interview data. These considerations are indicated in Krippendorff (2013) for the process of content analysis.

Specifically, the algorithm applied by the LDA identifies the most relevant words in the analyzed documents. In the present study, each interview was considered as a document. Using the topic-modeling process with LDA, we identified approximately 10 words for each document. These words were then used to form the names of topics in the data. This is a standard process in the use and development of LDA using the NLP framework. In the present study, the LDA process was computed with Python LDA 1.0.5 software.

Second, to complement the qualitative analysis outlined above with a quantitative assessment, we computed the key values of the identified topics. Keyness is a statistical indicator that measures the value, also known as the log-likelihood score (Rayson & Garside, 2000). This metric provides statistical meaning and makes it possible to measure the relevance of different topics in the same database or corpus. According to Duran, Hall, McCarthy, & McNamara (2010), the log-likelihood score of 3.8 or higher was reported to be statistically significant at $p < 0.05$. Therefore, the interview conversations were established as inputs phrases, and text documents were considered as sub-corpora of the original corpus. Statistical significance in this study was considered when $p < 0.05$ Drmota, Szpankowski, & Viswanathan (2012).

Furthermore, we used textual analysis computed in Python (Anand, Bochkay, & Chychyla, 2020). With this approach, it is possible to identify values in the form of insights using in-depth content analysis (Millstein, 2020). Specifically, the variables related to the weighted percentages/frequency of a keyword in the database composed of the set of interviews were studied (McHugh et al., 2020). In this way, the relevance of certain keywords was obtained (Auer, 2018). Based on the percentages of relevance achieved, we established parameters that casually explained the objectives of the present study (Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021b). This exploratory approach follows the indications of content analysis using the NLP framework.

An analysis of the main n-grams collected in the coded text of the interviews was also performed. In order to compute the n-grams analysis, we followed Wu and Su (1993) who argued that statistical analysis of the measure known as mutual information (MI) is justified when using textual analysis and n-grams. This indicator refers to the probability of co-occurrence of two variables that are correlated. Likewise, Bouma (2009) and Iyengar et al. (2012) used MI indicator between random variables X and Y. Of these, those with marginal probabilities and p(x) and p (y), and joint probabilities p (x, y), can be computed.

Results

According to the results of the systematic literature review, in the studies included in the dataset, we identified the main uses that governments make of AI. In this way, the interviews were developed based on the results of this methodological process. Furthermore, to complement the results obtained through our systematic literature review as indicated previously, we conducted interviews with informants who work or have worked in governments.

Regarding the major identified uses of AI by the governments, the main one is the continuous development of new models that increase the efficiency of the results (Chamola et al., 2020). This is a characteristic of AI, since the more the models that work with machine learning are trained, the greater the efficiency in terms of prediction of finance is, if the objects are focused on profitability. Likewise, the uses focused on decision making for process improvement and the evolution of management and governance practices were also remarkable (Skaug Sætra, 2020).

In this way, techniques are used to understand and optimize interactions with citizens (Androutsopoulou et al. (2019) through channels such as social networks (Saura, Palacios-Marqués, & Iturricha-Fernández, 2021; Silva et al., 2015), as well as information systems or data exchange platforms. Automation and the use of models and algorithms are being increasingly widespread in governments, as they, through new technologies linked to SI (chatbots, IoT, smart cities, among others), try to collect databases that can predict how society is organized, determine financial models, and improve the optimization of industries and cities (Silva et al., 2015; Zato et al., 2011).

References

1. Abou Elassad, Mousannif, Al Moatassime and Karkouch, 2020

   Z.E. Abou Elassad, H. Mousannif, H. Al Moatassime, A. Karkouch

   The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review

   Engineering Applications of Artificial Intelligence, 87 (2020), Article 103312, 10.1016/j.engappai.2019.103312

2. Acar, Englehardt and Narayanan, 2020

   G. Acar, S. Englehardt, A. Narayanan

   No boundaries: Data exfiltration by third parties embedded on web pages

   Proceedings on Privacy Enhancing Technologies, 2020 (4) (2020), pp. 220-238, 10.2478/popets-2020-0070

3. Agarwal and Dhar, 2014

R. Agarwal, V. Dhar

Big data, data science, and analytics: The opportunity and challenge for IS research

Information Systems Research, 25 (3) (2014), pp. 443-448, 10.1287/isre.2014.0546

4. Akter et al., 2019

   S. Akter, R. Bandara, U. Hani, S.F. Wamba, C. Foropon, T. Papadopoulos

   Analytics-based decision-making for service systems: A qualitative study and agenda for future research

   International Journal of Information Management, 48 (2019), pp. 85-95, 10.1016/j.ijinfomgt.2019.01.020

5. Akter and Wamba, 2016

   S. Akter, S.F. Wamba

   Big data analytics in E-commerce: A systematic review and agenda for future research

   Electronic Markets, 26 (2) (2016), pp. 173-194, 10.1017/S0963180114000589

6. Al-Mushayt, 2019

   O.S. Al-Mushayt

   Automating E-government services with artificial intelligence

   IEEE Access, 7 (2019), pp. 146821-146829, 10.1109/access.2019.2946204

7. Altman, Wood, O'Brien, Vadhan and Gasser, 2015

   M. Altman, A. Wood, D.R. O'Brien, S. Vadhan, U. Gasser

   Towards a modern approach to privacy-aware government data releases

   Berkeley Technology Law Journal, 30 (3) (2015), pp. 1967-2072

8. Anand, Bochkay, Chychyla and Leone, 2020

   V. Anand, K. Bochkay, R. Chychyla, A.J. Leone

   Using Python for Text Analysis in Accounting Research. Forthcoming, Foundations and Trends in Millstein, F. (2020). Natural language processing with python: natural language processing using NLTK

   Frank Millstein (2020)

9. Andrew and Baker, 2019

J. Andrew, M. Baker

The general data protection regulation in the age of surveillance capitalism

Journal of Business Ethics, 1-14 (2019), 10.1007/s10551-019-04239-z

10. Androutsopoulou, Karacapilidis, Loukis and Charalabidis, 2019

A. Androutsopoulou, N. Karacapilidis, E. Loukis, Y. Charalabidis

Transforming the communication between citizens and government through AI-guided chatbots

Government Information Quarterly, 36 (2) (2019), pp. 358-367, 10.1016/j.giq.2018.10.001

11. Ashok, Madan, Joha and Sivarajah, 2022

M. Ashok, R. Madan, A. Joha, U. Sivarajah

Ethical framework for artificial intelligence and digital technologies

International Journal of Information Management, 62 (2022), Article 102433, 10.1016/j.ijinfomgt.2021.102433

12. Auer, 2018

E.M.L. Auer

Detecting deceptive impression management behaviors in interviews using natural language processing

(2018)

13. Bacq, Janssen and Noël, 2019

S. Bacq, F. Janssen, C. Noël

What happens next? A qualitative study of founder succession in social enterprises

Journal of Small Business Management, 57 (3) (2019), pp. 820-844, 10.1111/jsbm.12326

14. Ballestar, Camiña, Díaz-Chao and Torrent-Sellens, 2021

M.T. Ballestar, E. Camiña, Á. Díaz-Chao, J. Torrent-Sellens

Productivity and employment effects of digital complementarities

Journal of Innovation and Knowledge, 6 (3) (2021), pp. 177-190, 10.1016/j.jik.2020.10.006

15. Belhadi et al., 2021

A. Belhadi, Y. Djenouri, G. Srivastava, D. Djenouri, J.C.W. Lin, G. Fortino

Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection

Information Fusion, 65 (2021), pp. 13-20, 10.1016/j.inffus.2020.08.003

16. Bem, 1995

D.J. Bem

Writing a review article for psychological bulletin

Psychological Bulletin, 118 (2) (1995), pp. 172-177, 10.1037/0033-2909.118.2.172

17. Benefo et al., 2022

E.O. Benefo, A. Tingler, M. White, J. Cover, L. Torres, C. Broussard, D. Patra

Ethical, legal, social, and economic (ELSE) implications of artificial intelligence at a global level: A scientometrics approach

AI Ethics, 1-16 (2022), 10.1007/s43681-021-00124-6

18. Bennett and Raab, 2020

C.J. Bennett, C.D. Raab

Revisiting the governance of privacy: Contemporary policy instruments in global perspective

Regulation & Governance, 14 (3) (2020), pp. 447-464, 10.1111/rego.12222

19. Biber, 2004

D. Biber

If you look at …: Lexical bundles in university teaching and textbooks

Applied Linguistics, 25 (3) (2004), pp. 371-405, 10.1093/applin/25.3.371

20. Biros, 2020

D. Biros

“the challenges of new information technology on security, privacy and ethics,” journal of the Midwest Association for Information Systems (JMWAIS), 2020, 2 Article, 1 (2020), 10.17705/3jmwa.000057

21. Blei, Ng, Jordan and Lafferty, 2003

D.M. Blei, A.Y. Ng, M.I. Jordan, J. Lafferty

Latent Dirichlet allocation

Journal of Machine Learning Research, 3 (2003), pp. 993-1022, 10.1162/jmlr.2003.3.4-5.993

22. Bouma, 2009

    G. Bouma

    Normalized (pointwise) mutual information in collocation extraction

    Proceedings of GSCL (2009), pp. 31-40
23. Bromberg, Charbonneau and Smith, 2020

    D.E. Bromberg, É. Charbonneau, A. Smith

    Public support for facial recognition via police body-worn cameras: Findings from a list experiment

    Government Information Quarterly, 37 (1) (2020), Article 101415, 10.1016/j.giq.2019.101415
24. Brynjolfsson & Mitchell, 2017

    E. Brynjolfsson, T. Mitchell

    What can machine learning do? Workforce implications

    Science, 358 (6370) (2017), pp. 1530-1534, 10.1126/science.aap8062
25. Cadwalladr and Graham-Harrison, 2018

    C. Cadwalladr, E. Graham-Harrison

    Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

    The Guardian, 17 (2018), p. 22

    Accessed on 27 January 2022 from:

    http://freestudio21.com/wp-content/uploads/2018/04/50-million-fb-profiles-harvested-by-cambridge-analitica.pdf
26. de Camargo Fiorini, Seles, Jabbour, Mariano and de Sousa Jabbour, 2018

    P. de Camargo Fiorini, B.M.R.P. Seles, C.J.C. Jabbour, E.B. Mariano, A.B.L. de Sousa Jabbour

    Management theory and big data literature: From a review to a research agenda

    International Journal of Information Management, 43 (2018), pp. 112-129, 10.1016/j.ijinfomgt.2018.07.005
27. Cao et al., 2014

L. Cao, T. Joachims, C. Wang, E. Gaussier, J. Li, Y. Ou, V.S. Subrahmanian

Behavior informatics: A new perspective

IEEE Intelligent Systems, 29 (4) (2014), pp. 62-80, 10.1109/MIS.2014.60

28. Cate, 2008

F.H. Cate

Government data mining: The need for a legal framework

Harv. CR-CLL Rev., 43 (2008), p. 435

29. Caudill and Murphy, 2000

E.M. Caudill, P.E. Murphy

Consumer online privacy: Legal and ethical issues

Journal of Public Policy & Marketing, 19 (1) (2000), pp. 7-19, 10.1509/jppm.19.1.7.16951

30. Chamola, Hassija, Gupta and Guizani, 2020

V. Chamola, V. Hassija, V. Gupta, M. Guizani

A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, Blockchain and 5G in managing its impact

IEEE Access, 1–1 (2020), 10.1109/ACCESS.2020.2992341

31. Chatterjee, 2019

S. Chatterjee

Impact of AI regulation on intention to use robots

International Journal of Intelligent Unmanned Systems, 8 (2) (2019), pp. 97-114, 10.1108/ijius-09-2019-0051

32. Chatterjee, 2020

S. Chatterjee

AI strategy of India: Policy framework, adoption challenges and actions for government

Transforming Government: People, Process and Policy (2020), 10.1108/tg-05-2019-0031 ahead-of-print(ahead-of-print)