Government Information Quarterly 40 (2023) 101877

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Trends and challenges of e-government chatbots: Advances in exploring open government data and citizen participation content

María E. Cortés-Cediel [a,*], Andrés Segura-Tinoco [b], Iván Cantador [b], Manuel Pedro Rodríguez Bolívar [c]

[a] Facultad de Ciencias Políticas y Sociología, Universidad Complutense de Madrid, Campus de Somosaguas, s/n, Pozuelo de Alarcón 28223, Spain
[b] Escuela Politécnica Superior, Universidad Autónoma de Madrid, Calle Francisco Tomás y Valiente, 11, Madrid 28049, Spain
[c] Facultad de Ciencias Económicas y Empresariales, Universidad de Granada, Paseo de Cartuja, 7, Granada 18011, Spain

ARTICLE INFO

Keywords:
Chatbots
Conversational agents
Digital government
E-government
E-participation
Open government data

ABSTRACT

In this paper, we propose a conceptual framework composed of a number of e-government, implementation and evaluation-oriented variables, with which we jointly analyze chatbots presented in the research literature and chatbots deployed as public services in Spain at national, regional and local levels. As a result of our holistic analysis, we identify and discuss current trends and challenges in the development and evaluation of chatbots in the public administration sector, such as focusing the use of the conversational agents on the search for government information, documents and services –leaving citizen consultation and collaboration aside–, and conducting preliminary evaluations of prototypes in limited studies, lacking experiments on deployed systems, with metrics beyond effectiveness and usability –e.g., metrics related to the generation of public values. Addressing some of the identified challenges, we build and evaluate two novel chatbots that present advances in the access to open government data and citizen participation content. Moreover, we come up with additional, potential research lines that may be considered in the future for a new generation of e-government chatbots.

1. Introduction

A key characteristic of Artificial Intelligence (AI) systems is their ability to learn from real-time, multi-modal inputs and to adjust their responses accordingly. Among other applications, public administrations have implemented AI-based technologies as a form of algorithmic bureaucracy, exploiting Natural Language Processing (NLP) techniques with the use of conversational agents or 'chatbots' as channels to provide information and services to citizens (Abbas, Følstad, & Bjørkli, 2023; Dhungel, Wessel, Zoubir, & Heine, 2021) and predictive analytics (Vogl, Seidelin, Ganesh, & Bright, 2020). Although the implementation of chatbots could introduce a challenge with unexpectedly positive results for public administrations (Androniceanu, 2023), it has allowed reducing service delivery costs, employee workloads, and waiting times of citizens for service assistance (Androutsopoulou, Karacapilidis, Loukis, & Charalabidis, 2019; Miner, Laranjo, & Kocaballi, 2020), as well as achieving more openness and transparency, by improving the accessibility to and reusability of government data (Kalampokis, Karacapilidis, Tsakalidis, & Tarabanis, 2023).

In brief, chatbots have come to disrupt some public processes and functions. Some examples of successful histories of chatbot implementations in public administrations include the French Ministry of Economy and Finance, which launched a conversational agent to provide access to human resource regulations for managers in the ministries of culture and social affairs (Afonasova, Panfilova, Galichkina, & Ślusarczyk, 2019), and the South Korea Central Government, which developed AI applications holding an international leadership position in digital governance and citizen participation (Szostak, 2022).

Nonetheless, up to now, chatbots in public administrations have been mainly focused on simple advice and information purposes (Abbas et al., 2023; Androutsopoulou et al., 2019; Cantador, Viejo-Tardío, Cortés-Cediel, & Rodríguez Bolívar, 2021; Makasi, Nili, Desouza, & Tate, 2020, 2022; Ramires Hernández, Valle-Cruz, & Méndez Mendoza, 2023; Song, 2022; Van Noordt & Misuraca, 2022), which has given place to confine these AI tools to written text messages (Rozenes & Cohen, 2022) and limit their potential for value creation into the public scope, often falling short of satisfying all the service delivery expectations (Makasi et al., 2020). Government chatbots have therefore been perceived useful

* Corresponding author.
E-mail address: mcorte04@ucm.es (M.E. Cortés-Cediel).

https://doi.org/10.1016/j.giq.2023.101877
Received 26 February 2023; Received in revised form 7 October 2023; Accepted 23 October 2023
Available online 4 November 2023

only for purposes of navigating available information and services, as well as for simple requests (Abbas et al., 2023). In this sense, for communication purposes, governments directly benefit citizens achieving uniformity in response quality and timeliness, and consequently enhance public trust, performance and effort expectancies for the chatbots use (Abbas et al., 2023; Aoki, 2020).

Although there has been recent research seeking to map AI applications in public sector entities –see e.g. (Alshehhi, Cheaitou, & Rashid, 2022; Ballester, 2021; Maragno, Tangi, Gastaldi, & Benedetti, 2021; Sandoval-Almazán & Millán-Vargas, 2023)–, as far as we are concerned, there is scarce literature on the chatbots use and characteristics in the public sector (Ramires Hernández et al., 2023). This is perhaps due to both the lack of effective adoption of chatbots within the public sector (Ramires Hernández et al., 2023) and the lack of comprehensive impact studies of chatbots in such setting (Van Noordt & Misuraca, 2022), resulting in a limited understanding of how chatbots have been deployed and evaluated within the public sector scope.

Indeed, as conversational and virtual agents rapidly advance in complexity, there are needs for more research regarding a wide number of aspects, such as the purposes, tasks and goals of the e-government chatbots, their target stakeholders and associated interactions and participation types, their implementation levels, the used communication interfaces and technologies, and the followed evaluation methodologies and metrics.

Under a holistic perspective, in this paper, we aim to shed light on trends and challenges concerning the above aspects for the analysis of chatbots in the public sector setting, seeking to provide not only guidance for public managers, politicians and practitioners to fully be aware and understand the public value creation produced by chatbots –assisting them in making decisions related to the development and exploitation of chatbots–, but also to researchers interested in the investigation of chatbots to identify fruitful topics and future work directions based on existing research gaps.

Concretely, our research seeks to address the following research questions: RQ1. Which are the trends and challenges of e-government chatbots identified in the research literature?; RQ2. Are there any gaps between the government e-chatbots studied in the literature and those that have been deployed by public administrations?; and RQ3. What could be novel and fundamental advances in the development of future e-government chatbots?

Data gathering methods to answer these questions include a combination of literature review (RQ1), the examination of real chatbot deployments in the Spanish public sector (RQ2), and the analysis of two empirical experiences conducted by the research team both in the use of chatbots to access open government data, and in the exploration of citizen-generated content of an e-participatory budgeting platform (RQ3). We believe that this research strategy represents an appropriate and holistic approach to achieve, taking into account such a broad spectrum in terms of coverage and scope.

Fig. 1 shows a schematic view of our work flow and paper structure. In Section 2, we will propose a conceptual framework to analyze e-government chatbots. The framework is built by considering existing research surveys on the topic (steps 1 and 2 in the figure), and is applied to analyze chatbots presented in research papers and chatbots deployed by public administrations. Hence, in Section 3, following the systematic PRISMA literature review methodology and applying our framework (steps 1 and 3), we will analyze prior research on e-government chatbots addressing RQ1; and in Section 4, taking the Spanish case as sample selection, we will apply the framework for an empirical analysis of chatbots currently deployed by Spanish public administrations addressing RQ2 (step 4). Afterward, targeting some of the challenges identified in the conducted analysis, in Section 5, we will present two novel e-government chatbots that were developed to support advanced information access –addressing RQ3–, and were evaluated by means of variables of our framework (step 5). Finally, in Section 6, we will bring the paper to an end with general conclusions and discussion.

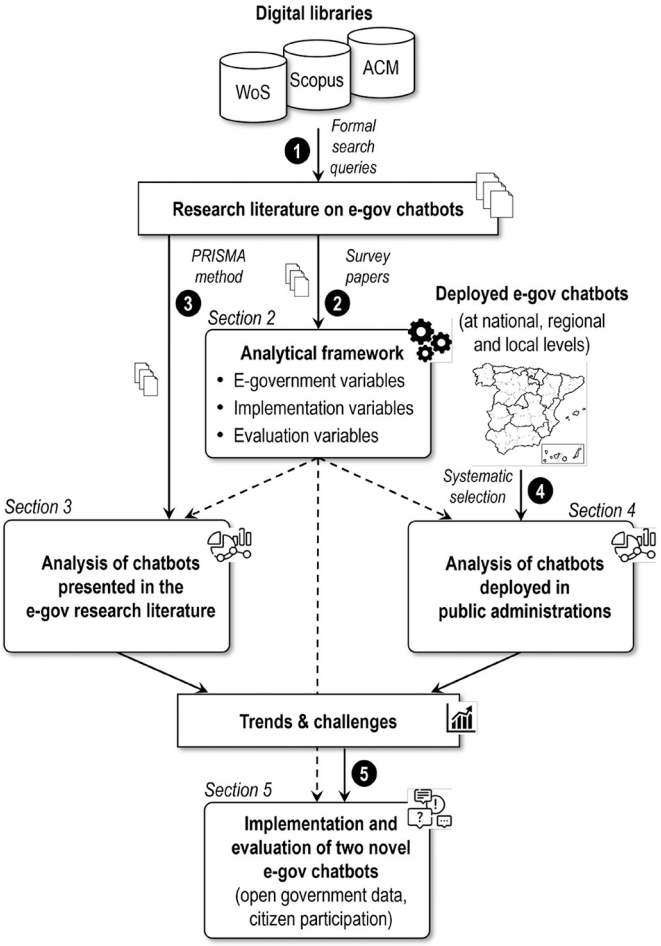

Fig. 1. Schematic view of our work flow and paper structure.

2. Analyzing e-government chatbots

Although in the research literature there are many papers that include sections describing related work on chatbots for certain applications and contexts of e-government (e.g., chatbots for accessing open government data), to the best of our knowledge, only three recent papers (Makasi, Nili, Desouza, & Tate, 2021; Nirala, Singh, & Purani, 2022; Ramires Hernández et al., 2023) are aimed to survey the state of the art on e-government chatbots; other review papers analyze several chatbots deployed in public administrations.

On an empirical sample of 14 papers (and 92 deployed chatbots), Makasi et al. (2021) proposed a typology for chatbots in public service delivery with three sophistication levels: service triaging, service information gathering and analysis, and service negotiation. For each of these levels of service complexity, the authors specified associated technical features (e.g., NLP and information retrieval, dialog management, and data storage) and capabilities (e.g., request coverage, query support, user-chatbot interaction, and response personalization). Nirala et al. (2022) also reviewed papers about chatbots in public administration, but restricted their analysis to technical aspects, such as the NLP techniques, dialog management technologies, and knowledge sources of the chatbots. Finally, considering a sample of 26 papers, Ramires Hernández et al. (2023) focused on the technologies and platforms currently used in e-government chatbots, and showed that Google Dialogflow[1] stands out (in 8 papers) as the predominant commercial

[1] Google Dialogflow chatbot development framework, https://cloud.google.com/dialogflow

technology to implement chatbots for public administration, and that the media where chatbots have mostly been integrated are websites and online social networks. Moreover, the authors mentioned that the main purpose of the surveyed chatbots is responding to citizens' questions. They, however, neither gave details on such matter, nor discussed other pursued purposes in the literature.

From the papers that analyze chatbots deployed in public administration, the most significant ones for our research purposes are those written by Van Noordt and Misuraca (2019) and Adnan, Hamdan, and Alareeni (2021). Van Noordt and Misuraca (2019) presented an exploratory study of 3 chatbots used in the public administrations of Latvia, Vienna and Bonn for answering citizens' frequently asked questions (FAQs) about administrative processes, aiming to reduce the resources employed in responding to the numerous citizens' calls and emails. More recently, Adnan et al. (2021) extended the scope of e-government chatbots deployed in Australia, Germany, North Carolina, Los Angeles, and Dubai for other functionalities beyond (frequently asked) question answering, such as information access, and online complaints, forms and payments.

Our research goes further to prior research, conducting a thorough, systematic review and comparison of e-government chatbots present in the research literature and chatbots deployed in public administrations; in particular, in Spanish administrations at national, regional and local levels. By contrast to Makasi et al. (2021) and Nirala et al. (2022), rather than focusing on algorithmic and technological aspects that are embedded into the design of the chatbots (which we do not entirely discard), we seek to analyze the sociotechnical lens underlying the chatbots use in the public sector setting. This sociotechnical analysis is built on the contribution of chatbots to create public value, which is not currently clear due to both the non-neutral effect of technologies (Vermaas, Kroes, Van de Poel, Franssen, & Houkes, 2022) and the non-understanding of how digital technologies can contribute to public value creation, especially when emerging technologies such as chatbots are being used (Panagiotopoulos, Klievink, & Cordella, 2019).

As stated by Benington and Moore (2010) and Dahl and Soss (2014), public value creation can be considered the outcome of a production process of public services pursued by public agencies to fulfill the collective goals. It means that public value is 'a social construct' (Morse, 2010) rooted in democratic theory (Albert & Passmore, 2008), making all public projects and initiatives to be citizen-centric and to create innovative forms of governance based on the concept of network governance (Rodríguez Bolívar, 2017).

According to Andrews (2019), debates about public value needs to take two major aspects into account: a) the value at the heart of the technology, including who can use data and for what purpose, and b) the deployment of the technological systems, which could be referred to the implementation aspects of chatbots that can also affect the level of service delivery sophistication and complexity (Makasi et al., 2021). Makasi et al. (2022) adds another aspect to be evaluated for public value creation in the chatbots' use for public administration, called "chatbot-mediated public service value dimensions," to ensure the benefits that the use of chatbots has for the public, including improved effectiveness and efficiency, greater sense of safety and sustainability, and higher trust in government (Makasi et al., 2020, 2022). This last aspect requires an evaluation against all the other public service value dimensions (Makasi et al., 2022).

Based on these sociotechnical lenses, our research focuses on the analysis of the public value contribution of chatbots examining the following three aspects: a) their purpose and potentialities for stakeholder interactions (e-government variables), partially considered by Van Noordt and Misuraca (2019); b) their implementation characteristics (implementation variables), considered by Ramires Hernández et al. (2023); and c) their accountability purposes (evaluation and metric variables), some of them preliminarily considered as benefits by Adnan et al. (2021). As noted previously, these attributes are the most relevant aspects for government researchers and practitioners to take decisions concerning the public value creation, based on the usefulness and potentiality of chatbots for public sector management and functionalities. In this way, our framework is composed of the following variables:

- **E-government-oriented variables**, which describe for what and for whom a chatbot has been developed, in the context of a public service. Specifically, the framework includes the *purpose* (e.g., searching for government information, accessing open government data, improving citizen participation), *task* and *goal* of a chatbot (e.g., reducing public administration costs, increasing government trust, promoting citizens' involvement), its *target stakeholders* (e.g., citizens, public administration operators, policymakers) and their *interactions*[2] (G2C, G2B, G2G), and the underlying *participation type* (information, consultation, collaboration).
- **Implementation-oriented variables**, which are associated to high-level technical aspects of a chatbot, discarding internal issues on how the chatbot has been implemented, such as natural language understanding and generation capabilities, and conversation intents, flow and management. We thus restrict our focus to the *implementation level* (proposal, prototype, deployed system), *communication interface* (text, buttons or menus, voice), *development technology* (e.g., Google Dialogflow, IBM Watson, Rasa) and *integration platform* (e.g., website, Twitter, Telegram) of the chatbot.
- **Evaluation-oriented variables**, which characterize how a chatbot has been evaluated: the followed *experimental method* (e.g., expert interview, user study), the considered *evaluation aspects* (e.g., effectiveness, efficiency, usability, privacy, trust), and the measured *metrics* (e.g., number of user actions, percentage of correct chatbot answers, opinion questionnaire responses).

As we shall show, the above variables can be extracted and analyzed from the chatbots surveyed in this work. The framework, nonetheless, may be open to further variables, especially if chatbot usage records and opinions from final users could be obtained after using the chatbots for a period of time.

As a result of the application of the framework on chatbots found in the research literature (section 3) and chatbots deployed in Spanish public administrations (section 4), our research will shed light on which are the trends and challenges of chatbots in the public sector setting (RQ1 and RQ2), and novel and fundamental advances in the development of future e-government chatbots (RQ3). The next sections are dedicated to these issues.

3. E-government chatbots in the research literature

This section is devoted to analyze the state of the art on e-government chatbots, focusing on the trends (subsection 3.2) and challenges (subsection 3.3) present in the research literature. Before, we describe the followed survey methodology and the obtained empirical sample for our analysis (subsection 3.1).

3.1. Survey methodology and empirical sample

To address RQ1 and subsequently RQ2, we conducted a systematic review of the research literature on e-government chatbots following the PRISMA methodology (Liberati et al., 2009). Fig. 2 shows the phases of our survey, which we describe next.

As done by Ramires Hernández et al. (2023), we retrieved the potential papers to review through a formal search query on digital libraries that index the publications of major journals and conferences; in particular, the Web of Science, Scopus and ACM libraries. Differently,

[2] We consider the types of interactions in e-governance: government-to-citizen (G2C), government-to-business (G2B) and government-to-government (G2G).

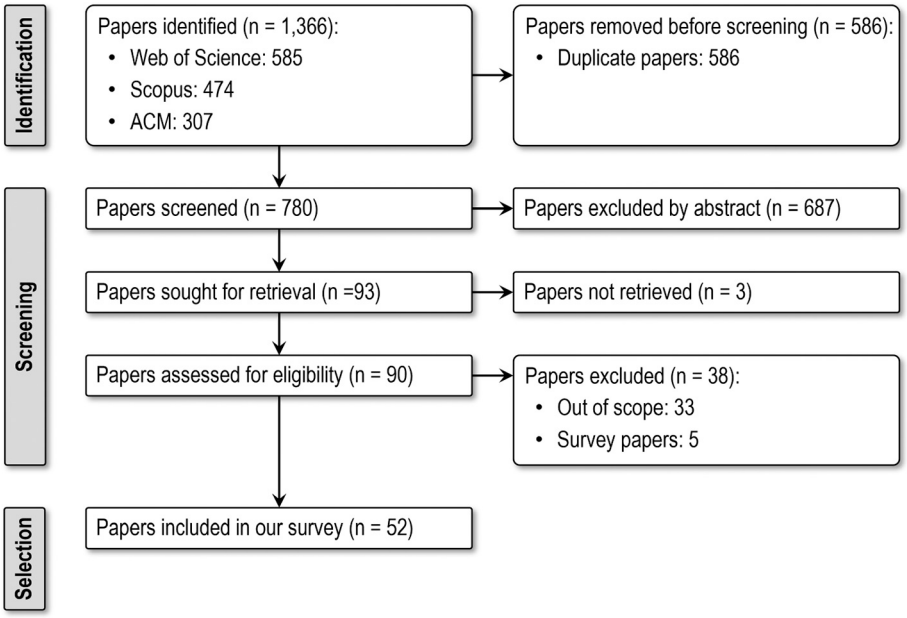

Fig. 2. Phases of our systematic survey following the PRISMA methodology.

our query included a large variety of keywords aiming to retrieve the maximum number of relevant publications, and our paper selection criteria did not limit to practical cases with which identifying the platforms and technologies used for the development of chatbots in public administration.

From the above libraries, our query retrieved those papers that include terms related to both chatbots and e-government. Table 1 shows the considered two sets of terms, TS1 and TS2, so that the retrieved papers should contain in their title, abstract or keywords some term from each set (i.e., column of the table). An asterisk * in a term entails a regular expression and means zero or more characters. Hence, for example, the expression "citiz*" is satisfied by the terms *citizen*, *citizens* and *citizenry*, among others.

The query was launched in January 2023, and retrieved a total of 1366 documents (including duplicated): 585 from the Web of Science, 474 from Scopus, and 307 from ACM. The documents were filtered or selected in two subsequent phases. In the first phase, we examined the abstracts of all papers to discard those that were clearly out of scope, mainly due to term ambiguities and work belonging to other disciplines, such as papers presenting chatbots on domains like education, health care, and tourism. In the second phase, we checked the content of the papers, considering for review those that present the proposal, implementation or evaluation of particular e-government chatbots, or propose or present categorizations or surveys of e-government chatbots. Overall, a final set of 52 relevant papers was selected up to 2023. From them, 33 papers described specific chatbots, and 29 papers addressed chatbot evaluation aspects. The number of journal and conference papers were 16 and 25, respectively. The journals and conferences mainly belonged to computer science (e.g., the Web and HCI conferences), government and public administration (e.g., the dg.o and ePart conferences, the Government Information Quarterly journal), and social sciences (e.g., the Journal of Information, Communication and Ethics in Society) fields. The remainder papers were workshop publications, except one, which was a book chapter. Although presenting ongoing work, we decided to include workshop papers since they may represent emerging, research trends. Regardless of the publication types, 78.8% of the selected papers were published in the last 5 years, which evidences the novelty and potential research opportunities of the topic.

To the best of our knowledge, our survey considers the greatest number of research papers on e-government chatbots, which were obtained through a formal search query that has a very large scope over several digital libraries. In comparison to previous, related surveys, ours is the first that aims to identify existing trends and reported challenges in the research literature. Moreover, differently to those surveys, which have focused on technologies and specific implementation aspects, ours considers sociotechnical features for the analysis of chatbots.

Table 1
Sets of terms used in the formal query that retrieves potential papers to analyze from digital libraries.

Term set 1 (TS1)	Term set 2 (TS2)
chatbot*, chatterbot*, conversational agent*, conversational assistan*, conversational bot*, conversational character*, conversational interface*, conversational system*, dialog* system*, dialog* bot*, dialog* agent*, dialog* interface*, dialog* character*, dialog* assistan*, digital assistan*, virtual assistan*, voice assistan*	government, digital government, e-gov*, egov*, public administrat*, public sector*, public service*, open government*, open data, open administrat* data, open government data, citizen* participation, e-participati*, eparticipati*, civic*, e-consult*, econsult*, e-vot*, evot*

3.2. Trends in the surveyed chatbots

To analyze the e-government chatbots compiled from the surveyed papers, we consider the variables of our framework presented in section 2. In appendices A and B, we provide a number of tables characterizing and classifying all the papers according to such variables. In the next subsections, we summarize and discuss the tables.

3.2.1. E-government-oriented trends

Aiming to identify the trends that have been followed so far in the literature on e-government chatbots, one may first ask the question: What are the purposes of the chatbots created for the public sector setting according to prior research?

From our survey, we have observed five principal purposes, namely 1) searching for government information, documents and services; 2) supporting the access to open government data; 3) providing public services; 4) improving citizen participation; and 5) facilitating the communication between stakeholders.

At the information level of participation, the most recurrent purpose of chatbots addressed in the literature is **searching for government**

information, documents and services, for which citizens have been the most targeted stakeholders (i.e., following a G2C interaction). In this context, we can distinguish several tasks and goals, with different degrees of complexity. The simplest one is using a chatbot for exploring Frequently Asked Questions, FAQs, related to administrative services (De Lacerda & Aguiar, 2019; Hasan, Rizvi, Jain, & Huria, 2021; Lommatzsch, 2018; Lommatzsch & Katins, 2019). More complex tasks include employing traditional information retrieval techniques to perform keyword-based queries for retrieving indexed texts, documents and services (Atreja et al., 2018; Boden, Fischer, Herbig, & Spierling, 2006; Valtolina, Barricelli, Gaetano, & Diliberto, 2018). For instance, Gatius, González, Militello, and Hernández (2006), Gatius, González, & Comelles, 2007) propose a chatbot (and voice agent) that searches for public services, by matching terms expressed in user queries with ontology concepts that belong to semantic descriptions of the services. Moreover, there are chatbots that make use of question answering (Q&A) systems for obtaining concrete answers to input questions expressed in natural language (Acer, Van den Broeck, & Kawsar, 2019b; Kucherbaev, Psyllidis, & Bozzon, 2017; Lockett et al., 2019; Mahapatra, Sharma, Trivedi, & Aman, 2012), from connected government data sources (e.g., municipal and regional databases).

Apart from these citizen-oriented applications, chatbots have been proposed for G2G interactions, to assist public administration operators (Bagnasco, Cappelli, & Magnini, 2000) (allowing them to access several data sources during a session with a citizen) and policy-makers (Kucherbaev et al., 2017) in information retrieval and analysis tasks (allowing them to request particular data as decision support). Moreover, providing G2B interactions, a chatbot has been developed in South Korea to provide business consulting and support for foreign enterprises (Thai & Huh, 2021); the chatbot is capable of exploring investment opportunities, according to government policies, laws and taxes, among other issues.

Also at the information level, a second principal purpose of chatbots is **supporting the access to open government data**. In this case, we identify a first, simple task: searching for data collections that are of interest for the user, according to an input topic/location (Keyner, Savenkov, & Vakulenko, 2019) or query (Neumaier, Savenkov, & Vakulenko, 2017). In general, these chatbots make use of information retrieval techniques, and provide as results links that point to the external collections, which have to be explored independently to the chatbot. More advanced work has investigated the use of chatbots for a second, much more complex task: exploring the content within data collections. In this case, we observe a variety of approaches: launching a limited set of SQL queries to relational databases created from open data collections (Porreca, Leotta, Mecella, & Catarci, 2017; Porreca, Leotta, Mecella, Vassos, & Catarci, 2018), exploring structured knowledge bases (built with linked open data) via predefined SPARQL queries (Anelli, Di Noia, Di Sciascio, & Ragone, 2019; Ronzhin et al., 2019), adopting application programming interfaces, APIs, as mechanisms to facilitate the access to open data collections (Sánchez-Nielsen, Morales, Mendo, & Chávez-Gutiérrez, 2021), and dynamically building and launching SQL queries through natural language conversations (Cantador et al., 2021).

As a third purpose, chatbots have been developed for **providing public services** distinct to information search and open data access, such as applying for an identity card or passport, marriage, and registration or change of home address. At information and consultation levels, Griol and García-Jiménez (2012) present a voice agent that offers a variety of City Council services in addition to access to city information, such as making administrative procedures, surveys, suggestions and complaints, and transferring to a public administration operator. Valverde (2019) present a chatbot that permits accessing to a number of business and citizen services, such as changing and renewing the citizen card, and revalidating the driving license. Androutsopoulou et al. (2019) present a rich, prototype chatbot that is built upon NLP, machine learning and data mining technologies, and allows performing public administration tasks using government data of various forms. Finally, Gerontas et al. (2022) present a chatbot that serves as an entry point to multiple public services that are semantically described and personalized via the Core Public Service Vocabulary Application Profile (CPSV-AP), a European standard to facilitate public service catalogs and interoperability.

Besides, targeting stakeholders distinct to citizens, we found the proposal of a chatbot to be used as human-computer interactive interface with decision support systems. Providing G2G interactions, the chatbot is envisioned as a specialized advisor for government managers of a (virtual) town, capable to suggest strategic decisions considering the current status and uncertainty conditions of the town (Augello, Pilato, & Gaglio, 2009, 2011).

A fourth principal purpose of chatbots is **improving citizen participation**. In this case, chatbots have been considered for both information and consultation levels. At the information level, researchers have proposed chatbots as credible sources of political information for government elections, aimed to promote voting (Martin, Allagha, & Misnikov, 2021) and to give voting advice according to personal preferences (Kamoen & Liebrecht, 2022). At the consultation level, we identify a first set of chatbots that allow citizens reporting information to government and municipal authorities; for instance, incidents during elections (Meng & Khelladi, 2017), civic issues and complaints (Atreja et al., 2018), and urban issues, e.g., potholes on roads (Kucherbaev et al., 2017). We observe a second set of chatbots used as e-participation tools to promote the process and documentation of citizens' idea generation (Tavanapour, Poser, & Bittner, 2019), moderate online discussions and develop argumentation reasoning (Haqbeen, Sahab, Ito, & Rizzi, 2021), support young people to participate in civic activities (Väänänen, Hiltunen, Varsaluoma, & Pietilä, 2020), and facilitate argument-driven access to citizen generated content in e-participatory budgeting platforms (Segura-Tinoco, Holgado-Sánchez, Cantador, Cortés-Cediel, & Rodríguez Bolívar, 2022).

Finally, a less extended, but highly relevant and promising purpose of chatbots in e-government is **facilitating the communication between stakeholders**. As a G2C interaction case, in (Kucherbaev et al., 2017), chatbots are envisioned as mediators in conversations between citizens and municipality employees for attending particular information or service needs (at information level), and between citizens and policy-makers for providing feedback about new policies and regulations (at consultation level). As a C2C interaction case, in (Portela, 2021), a chatbot is presented to support the communication with others in citizen science projects (at collaboration level).

In addition to the above purposes, from the retrieved papers, we also identified particular e-government applications of chatbots: attendance and management in pandemics, mostly COVID-19 –e.g., (Miner et al., 2020; Tanoue et al., 2020),– and in crisis and disaster situations –e.g., (Ahmady & Uchida, 2020; Piccolo, Roberts, Iosif, & Alani, 2018)–, health care –e.g., (Ávila et al., 2019)–, education –e.g., (Guy de Andrade et al., 2020)–, job –e.g., (Bellini et al., 2020)–, law –e.g., (Firdaus, Saputra, & Suprianto, 2020)–, immigration –e.g., (Drydakis, 2021)–, tourism –e.g., (Massai, Nesi, & Pantaleo, 2019)– and public media –e.g., (Ford & Hutchinson, 2019; Massai et al., 2019). We omit research work on these domains, since it should be reviewed thoroughly with specific searches on specialized fields, which are out of the scope of this study.

3.2.2. *Implementation-oriented trends*

With respect to e-government chatbot implementation issues, our survey reveals that, in terms of implementation level, the majority of published work on the topic deals with proposals (15% of the analyzed papers) and prototypes (67%) of chatbots. Only a few works (18%) present conversational agents that were deployed in real systems and tested at scale (De Lacerda & Aguiar, 2019; Lommatzsch & Katins, 2019; Meng & Khelladi, 2017; Portela, 2021; Thai & Huh, 2021). We omit here those researches aimed to evaluate chatbots currently used by public administrations, and commonly developed by third-party companies (see subsection 3.2.3).

In general, in the research literature, textual conversations represent the priority communication channel between users and chatbots; it is considered in 82% of the surveyed papers –e.g., (Haqbeen et al., 2021; Portela, 2021; Thai & Huh, 2021). In certain cases (30%), dialogs based on menus and buttons used in isolation or in combination with simple text commands have been considered as well –e.g., (Gerontas et al., 2022; Kamoen & Liebrecht, 2022; Sánchez-Nielsen et al., 2021). There are a few papers that cite the possibility of voice-based assistance. However, nowadays, this is not a critical point, since commercial chatbot technologies, such as Google DialogFlow, already include effective speech-to-text interfaces.

As one may expect, the chatbot types and technologies considered in the research literature reflect the evolution and recent, impressive advances of the NLP field, for both language understanding and conversation management. Thus, in the surveyed papers, we can identify a trend that has shifted from ad hoc implementations of chatbots using limited keyword matching- and rule-based methods, e.g., using the AIML standard (Augello et al., 2011; Boden et al., 2006; Mahapatra et al., 2012; Meng & Khelladi, 2017), to advanced, large-scale chatbot implementations based on machine (neural network) learning NLP models (Lommatzsch, 2018; Lommatzsch & Katins, 2019), in most cases using well-known, easy-to-use technologies of big companies, such as Rasa (De Lacerda & Aguiar, 2019; Gerontas et al., 2022; Keyner et al., 2019; Ronzhin et al., 2019), DialogFlow from Google (Anelli et al., 2019; Cantador et al., 2021; Hasan et al., 2021; Sánchez-Nielsen et al., 2021; Segura-Tinoco et al., 2022; Valtolina et al., 2018), Azure Bot from Microsoft (Neumaier et al., 2017; Valverde, 2019), and Watson from IBM (Atreja et al., 2018; Porreca et al., 2018).

This last trend is related with the digital platforms in which chatbots have been integrated: from public administration systems –e.g., (Bagnasco et al., 2000; De Lacerda & Aguiar, 2019; Lommatzsch, 2018)– and web portals –e.g., (Haqbeen et al., 2021; Valverde, 2019)–, to instant messaging applications –such as Telegram (Cantador et al., 2021; Segura-Tinoco et al., 2022) and WhatsApp (Gerontas et al., 2022)– and online social networks –such as Facebook (Neumaier et al., 2017) and Twitter (Meng & Khelladi, 2017).

3.2.3. Evaluation-oriented trends

In our review of the literature related to the evaluation of e-government chatbots, we were able to identify three research trends, addressing topics of interest for different stakeholders.

The first trend (43% of the surveyed evaluation papers) focuses on major **concerns about e-government chatbots**, from the point of view of public administration actors (Petriv, Erlenheim, Tsap, Pappel, & Draheim, 2019), development teams (Baldauf & Zimmermann, 2020; Petriv et al., 2019), and end users (Baldauf, Zimmermann, & Pedron, 2021). Usually, based on data collected in interviews, user questionnaires, and expert surveys, these concerns are studied in order to be formalized and empirically confirmed (or discarded). In most cases, these studies focus on desired characteristics –such as, *acceptability* (Akkaya & Krcmar, 2019; Baldauf et al., 2021; Nowakowska-Grunt, Dziadkiewicz, Olejniczak-Szuster, & Starostka-Patyk, 2021; Wright, 2021), *trust* (Aoki, 2020; Pal, Arpnikanondt, Razzaque, & Funilkul, 2020), *privacy* (Arifin & Lennerfors, 2021; Baldauf & Zimmermann, 2020; Pal et al., 2020), *user satisfaction* (Tisland, Sodefjed, Vassilakopoulou, & Pappas, 2022), and *user engagement* (Portela, 2021), among others–, or on measuring the adoption intent (Akkaya & Krcmar, 2019; Kuberkar & Singhal, 2020) of a potential chatbot.

The second trend (21% of the papers) comes from a set of papers presenting **studies on user-chatbot interactions**, with respect to the completion of particular e-government tasks. Usually, the data used in these papers are collected from user logs files (as done e.g. in (Cantador et al., 2021; Segura-Tinoco et al., 2022; Valverde, 2019), but other primary sources of information are user questionnaires and expert surveys –e.g., (Baldauf & Zimmermann, 2020; Chohan, Hu, Khan, Pasha, & Sheikh, 2021; Kuberkar & Singhal, 2020). In this trend, the most commonly used task-dependent metrics are *effectiveness* (Acer, Van den Broeck, & Kawsar, 2019a; Nowakowska-Grunt et al., 2021; Valverde, 2019), *efficiency* (Acer et al., 2019a; Baldauf & Zimmermann, 2020; Valtolina et al., 2018) and *acceptability* (Baldauf et al., 2021; Chohan et al., 2021; Kuberkar & Singhal, 2020; Vassilakopoulou, Haug, Salvesen, & Pappas, 2022). However, *professionalism* (Acer et al., 2019a; Baldauf & Zimmermann, 2020) and *user satisfaction* (Cantador et al., 2021; Kuberkar & Singhal, 2020) are also considered.

Finally, the third trend (36% of the papers) consists of **evaluations of deployed chatbots** in online public administration environments. In this case, only Simonsen, Steinstø, Verne, and Bratteteig (2020) rely on user logs to compute evaluation metrics. The remainder works use expert interviews (Petriv et al., 2019), questionnaires or surveys (Griol & García-Jiménez, 2012; Valtolina et al., 2018; Vassilakopoulou et al., 2022) –e.g., the User Experience Questionnaire, UEQ (Rauschenberger, Schrepp, Cota, Olschner, & Thomaschewski, 2013)–, and laboratory studies (Federici et al., 2021) as data sources. In all cases, researchers mainly use *effectiveness*, and *acceptability* and *usability* metrics to respectively evaluate the performance and quality of the chatbots.

3.3. Reported challenges

From the analysis outlined in the previous subsections, in the research literature, we first observe a predominance of chatbots that have not been deployed and evaluated in real public administration applications, at scale. Moreover, although the majority of chatbots work at the information level of participation, there is a significant number of prototypes that address the consultation level, allowing stakeholders (mostly citizens) to provide information to government entities. With the aim at both developing advanced conversational agents at the consultation and collaboration levels and conducting more exhaustive evaluations, information sharing from government entities is a big problem that should be solved (Sandoval-Almazán & Gutiérrez-Alonso, 2009).

With respect to the access to open government data, in general, chatbots are limited to searching for data collections given certain user needs, commonly expressed as keyword-based queries. In this sense, there is the necessity to research novel conversational agents that allow digging into the content of such collections, preferably through sophisticated, formal queries (Cantador et al., 2021). Moreover, the chatbot provision of summaries and analyses of the retrieved data would certainly be very valuable for decision-making tasks.

Regarding citizen participation, researchers have identified several challenges for chatbots, such as filtering or matching similar ideas gathered in e-participation platforms (Tavanapour et al., 2019), aggregating external sources to complement citizen reports (Meng & Khelladi, 2017), and implementing functionalities oriented to mediate and facilitate the collaboration among stakeholders (Segura-Tinoco et al., 2022).

Apart from these challenges, general open issues concerning *conversational features*, *user-related aspects* and *evaluation issues* have been mentioned in the literature on e-government chatbots. Related to the conversation maintained with the chatbots, the provision of more flexibility on the natural language formulation of user utterances, and the generation of user-friendly (and even colloquial) responses from the chatbots, are desired features (Sandoval-Almazán & Gutiérrez-Alonso, 2009; Segura-Tinoco et al., 2022). As advanced functionalities, some researchers have claimed the need for effectively managing the context, coherence and flow of a conversation (Portela, 2021), and even dealing with follow-up questions that are answered with respect to past conversations (Lommatzsch, 2018). As more technical aspects, it has been evidenced the need for handling spelling errors (Atreja et al., 2018; Simonsen et al., 2020), having the capacity of summarization, repair, repeat or paraphrase (Portela, 2021), and enabling additional interfaces, such as a map-based (Atreja et al., 2018).

With respect to user-related aspects, some authors have expressed that e-government chatbots should provide affective, emotional and

social cues (Portela, 2021), and maintain a natural and friendly conversation regarding tone and sensitivity, showing a pleasant personality, being empathetic and non-judgmental, and developing a personal relationship with the user, e.g., remembering her name (Tavanapour et al., 2019). Other authors have recognized the importance of providing chatbot with capabilities to recognize the users' preferences, in order to proactively make personalized recommendations of relevant content, mitigating situations of information overload (Portela, 2021; Segura-Tinoco et al., 2022; Simonsen et al., 2020). There even exists the proposal of applying gamification strategies by chatbots to motivate and engage the users, e.g., in citizen participation tasks (Kucherbaev et al., 2017).

Regarding evaluation issues, we have seen that there is no consensus on the definition of the used metrics which, on the other hand, mainly focus on system effectiveness and efficiency, and global user acceptability and satisfaction. In this context, public values-based concerns are rarely considered in the development of e-government chatbots. Being transparent, providing trustfulness, and ensuring and informing about privacy are some of the mentioned desired features in that respect (Portela, 2021). Important aspects, such as accountability (explainability), fairness and legitimacy, have recently been taken into account in the evaluation of chatbots (Cantador et al., 2021; Segura-Tinoco et al., 2022), but not in their design and implementation.

4. E-government chatbots in deployed applications

In this section, we empirically analyze e-government chatbots currently used by public administration, aiming to identify general trends (subsection 4.2) and challenges (subsection 4.3), and to compare them with those present in the research literature and analyzed in section 3. Before, we describe the followed methodology to select the studied chatbots, and the resultant sample of our analysis (subsection 4.1).

4.1. Survey methodology and empirical sample

To address RQ2, we empirically analyze conversational agents deployed in public administrations. With the aim of complementing the analysis presented in the previous section, and differently to previous studies on deployed chatbots in the research literature –e.g., (Adnan et al., 2021; Van Noordt & Misuraca, 2019)–, which only consider a very limited number of chatbots as case studies, we conducted a systematic search of the e-government chatbots currently deployed into all levels of the Spanish public administration.

Specifically, we searched for chatbots accessible via official websites and open data portals at different administrative division levels, namely national, regional and local levels. Thus, we considered how the government power is territorially articulated in Spain. First, we searched for chatbots used by the central, national government entities; in particular, the 22 Ministries that are currently part of the Government of Spain, and by major Spanish public institutions,[3] such as the Congress of Deputies, the Court of Auditors, and the Ombudsman. We also considered certain additional e-administration structures at national level, due to their large scope and relevance.

At the regional level, we searched for chatbots deployed by the 17 Regional Governments and the 2 Autonomous Cities (Ceuta and Melilla) that integrate the main regional level of public administration in Spain. Finally, at the local level, we considered chatbots from the set of municipalities (88 cities) that belong to the Spanish Network of Smart Cities (RECI).[4] These cities are those with a high technological level involved in all their actions and management systems, which fits well with the aim of our research.

Despite the previous criteria, some chatbots deployed by Spanish public administrations have not been included into our sample selection. First, we may be omitting some private chatbots that are being used internally in administrations. Second, we discard chatbots at local level that are deployed by city councils of medium-size municipalities because they have similar characteristics and functionalities than those of our sample. Finally, we also exclude some chatbots that are not currently operational, such as those intended to answer questions for citizens during the COVID-19 pandemic. Besides this, we also note that, at the time of writing, there are ongoing projects of Spanish public administrations that are aimed to develop new conversational systems within the national AI strategic plan.

After reviewing all potential study cases, from a set of 153 public administration entities, we identified a total of 25 chatbots deployed at the different public administration levels (5 chatbots at the national level, 4 at the regional level, and 14 at the local level) –cf. tables C1 and C2 of appendix C. These chatbots are analyzed in subsequent subsections.

The exhaustive research survey we have conducted allows us to claim that our study is the only one that follows a systematic methodology to search, select and analyze chatbots deployed by public administrations with no academic purposes. This, together with the distinction of national, regional and local administrative levels, enable the reproducibility of the study and its extension to other countries and specific public sectors.

4.2. Trends in the surveyed chatbots

Similar to what was done in our research literature survey, we apply the framework[5] proposed in section 2 to perform an analysis of the chatbots deployed in the Spanish public administration. In appendix C, we provide table C3, which characterizes and classifies the chatbots with respect to the framework variables. Next, we summarize and discuss the table.

4.2.1. E-government-oriented trends

One of the principal findings in our study is the fact that all the analyzed chatbots operate at the information level of participation, and follow a G2C interaction. This is in accordance with the research literature, which, by contrast, includes some examples of chatbots that operate at consultation level, and follow G2B and G2G interactions.

Although the chatbots are aimed to offer information to citizens, they have relevant particularities, as shown next. Most of the chatbots (21 out of the 25 analyzed cases) allow **searching for government information**. In some cases, such information is related to the provision of public services. For example, the *AVIVA* virtual assistant –deployed by the Spanish Ministry of Finance and Civil Service– guides citizens in tax processing and payment, and the *ISSA* agent –belonging to the Spanish Ministry of Inclusion, Social Security and Migration– helps citizens to apply for public benefits. In general, the information provided by the chatbots is diverse on a variety of types of communication and procedures. For example, the *Defensor del Pueblo* agent –running upon the Ombudsman web portal– guides citizens in filing formal complaints with different agencies, in cases where they feel that their rights have been damaged. The complaints can be processed through the conversational agent itself. Besides, some of the chatbots belong to specific and specialized domains. This is the case of agents that offer information about health care (*CoActuem*), tourism (*Goio*, *Turismo de Fuengirola*, and *CarnavalSC23*), and environment and water management (*AIRE* and

[3] Major Spanish public administrations, https://www.lamoncloa.gob.es/lang/en/espana/spanishinstitutions/

[4] Spanish network of smart cities, https://reddeciudadesinteligentes.es/

[5] We note that evaluation-oriented variables of the framework cannot be applied, since, to the best of our knowledge, evaluations of the deployed chatbots have not been reported by the Spanish public administrations.

Dra. Margon).

We found out 2 chatbots aimed at **supporting the access to open government data**: *Xatkit* and *AOD Chat*, made available by the Government of Spain and the Regional Government of Aragon, respectively. Through these chatbots, citizens can search for open data collections belonging to multiple domains. However, they are not able to dig into the content of such collections.

The remainder 2 chatbots, which also provide government information, have the additional purpose of **facilitating the communication between stakeholders**. *Clara*, the chatbot deployed upon the 'Decide Madrid'[6] participatory budgeting web platform, offers information on citizen participation procedures. The *WebChat* chatbot –implemented at the national level– redirects citizens to other information channels when their queries are not satisfied by the chatbot.

In addition, according to the e-government oriented variables, we observe that the chatbots are generally valuable for distinct stakeholders. However, it should be noted that in most of the chatbots implemented at the local level, they cover the information given to citizens transversally, acting as PSC or points of single contact (e.g., the *Ajuntament de Mataró Bot*). This issue is related to the goal of reducing costs to the public administration, by facilitating the automation of administrative processes.

4.2.2. Implementation-oriented trends

Next, we analyze some implementation aspects of the considered deployed chatbots. First, all the systems follow a simple design, and almost all of them are integrated on websites, except the cases of *CoActuem* and *CarnavalSC23*, which are accessible via Telegram and WhatsApp instant messaging applications, respectively.

Second, most of the analyzed chatbots have a text-based interface. In contrast, only one chatbot of the sample is based on text and images (*AIR*), and only one supports text and voice (*Turismo de Fuengirola*). Some of the text-based chatbots also use button menus (e.g., *María la Cigüeña* and *Arminda*). These menus often have a limited set of questions or answers that citizens must stick into their needs and queries (e.g., *La Abuela Elvira*), reducing the information exploration capabilities of the chatbots.

Moreover, we observe that almost all the chatbots do not allow downloading items. The exceptions are *La Abuela Elvira*, *CoActuem* and *Dra. Margon*. Through these chatbots, citizens can obtain government documentation, such as administrative regulations and public data of interest.

4.3. Identified challenges

Once the set of chatbots implemented in the Spanish public administration has been analyzed, we claim several findings. First, when comparing the three levels in which power is distributed territorially in Spain (central, regional and local), there are some particularities in the chatbots belonging to each level. On the one hand, the local level has associated a greater number of implemented chatbots (19) compared to those implemented at the central (3 chatbots) and regional (3 chatbots) levels. Furthermore, there are differences in the type of information with which chatbots interact with citizens. Specifically, most of the chatbots implemented at the local level (12 chatbots) offer transversal information and help on any issue that citizens need. However, in the local level, a reduced number of chatbots are specialized in specific topics, such as *Goio* in Canary Islands, which focuses on the tourism domain.

By contrast, most of the chatbots implemented at the central level of the public administration in Spain have a sectoral perspective in relation to the information they provide to citizens. Hence, for example, we find *AVIVA* and *Issa* chatbots, which are focused on helping citizens in the management of public taxes and benefits, respectively. In this sense, it would be interesting for chatbots to integrate other types of related information that help citizens to centralize assistance, with any questions they may have, increasing thus efficiency and effectiveness in management tasks.

Related to design aspects and functionalities, the identification of a gender for the analyzed chatbots, either by their name or by the avatar attributed to them, is striking. In particular, 60% of the chatbots are endowed with a neutral gender that does not correspond to the feminine and masculine genders (15/25), 32% (8/25) have female names or female-looking avatars, and only 8% (2/25) correspond to the male gender.

In the research literature, there is a tendency to prefer female names and female-looking avatars for chatbots related to customer services and sales (Feine, Gnewuch, Morana, & Maedche, 2020; Zogaj, Mähner, Yang, & Tscheulin, 2023), which implies the existence of certain gender bias in the chatbots design (Feine et al., 2020). Differently, in the case of the analyzed deployed chatbots, it is observed that there is a majority has a neutral gender, trying to highlight an aspect more related to the concept of robot. This may influence on the extent in which relations between public administration and citizens are strengthened. According to authors such as Zogaj et al. (2023), it is necessary the fair and unbiased development of features that contribute to the perceived humanization of nonhuman entities, with the aim of increasing trust and confidence in the interactions between humans and machines.

In addition, there are other aspects related to the design of the deployed chatbots that are worth to be pointed out. From the analysis reported in the previous subsections, we observe that, in general, the chatbots that currently exist in the Spanish public administration are not personalized and sophisticated tools. On the contrary, they are simple technologies with basic and limited features and functionalities. Specifically, most of them have been designed to provide information of interest on regulations and events, and some on the processing of services.

In our empirical sample, the chatbots do not enable or improve participation scenarios. Thus, from the point of view of the participation levels, one of the first identified challenges is the need for developing chatbots to promote citizen involvement at the levels of consultation and collaboration, for both decision- and policy-making. In this context, it could be interesting to create chatbots aimed to favor communication between stakeholders. This, among other things, may entail expanding services to G2B or G2G interactions.

Another improvement that could be addressed in the Spanish public administration sector is the use of chatbots able to provide complex public services as those found in the surveyed papers, which take advantage of current conversational agents technologies, such as advanced access to open government data Cantador et al. (2021).

Finally, it is convenient to provide chatbots with appropriate accessibility and usability (e.g., ease of use). In this sense, among other issues, the incorporation of features like the use of voice, images and multimedia (i.e., audios, videos and maps) into the user-chatbot communication could be considered.

5. Advanced e-government chatbots

Addressing RQ3, in this section, we present two chatbots built upon Google Dialogflow that enable advanced access to open government data (subsection 5.1) and citizen participation content (subsection 5.2), aimed at fostering transparency, accountability and public value creation. The chatbots and their evaluation were initially presented and detailed in (Cantador et al., 2021) and (Segura-Tinoco et al., 2022), making use of some of the variables of the conceptual framework proposed in section 2. Therefore, we summarize them, highlighting their novelties with respect to the state of the art, and reporting some evaluation results.

[6] Decide Madrid e-participatory budgeting platform, https://decide.madrid.es

5.1. A chatbot for exploring open government data

As seen in our survey of the research literature (section 3) and the Spanish public administration (section 4), in general, the chatbots of open government data portals allow searching for data collections, but not exploring the data of the collections. As a novel contribution, our first chatbot enables the access to data through complex, formal SQL[7] queries that are easily and dynamically built through a natural language conversation.

The chatbot operates on SQL relational databases and Apache Lucene[8] indices created automatically from open data collections. In conducted experiments, we used collections from the open data portal of Madrid City Council.[9] The catalog of the portal contains more than 570 data collections from different public sectors (e.g., education, environment, public transport, etc.), published in different formats (mainly CSV[10] files and Microsoft Excel spreadsheets) and with different update frequencies (such as daily, weekly, monthly, etc.). Among other metadata, each collection has a title and a description of its inner structure, which we used to build the associated database and index.

5.1.1. Conversation intents

The chatbot conversation flow is composed of nine intents that represent different user needs (purposes or goals). Each intent is independent of the rest of intents, and is considered after addressing another particular intent. An intent is triggered if the user enters a natural language sentence that satisfies a certain pattern, which is automatically learned by the NLP module of Google DialogFlow from input, representative sentences provided in advance by the chatbot developers.

The chatbot has a set of intents that allow the user to list, search and select available collections. Examples of identified sentences are "What collections are available?," "Search for a collection…," and "I would like a collection about…" When a user has selected a particular collection (establishing the FROM component of a SQL query), the chatbot allows exploring the fields (identifiers and names) of the associated table in the database, and launching a query against the table with or without filters.

In both types of queries, the user selects the fields of the table the user is interested in (i.e., the SELECT elements of the SQL query). In a query without filters, there are no criteria to restrict the records of the table to retrieve. Differently, in a query with filters, the user is able to establish criteria to be satisfied by the retrieved records on the values of some of their fields (i.e., the WHERE clauses of the SQL query). These criteria are of the form "field operator value", where operator is a relational operator (e.g., = for *equal to*, > for *greater than*, and!= for *not equal to*).

Fig. 3 shows three screenshots with fragments of a conversation maintained in our chatbot. In the left fragment, the user asks for collections. The chatbot responds by asking the user for the terms with which performing the search. Afterward, it shows the titles of the retrieved collections. In the middle fragment, the user states a query with filters to be performed (to a previously selected table). Then, the chatbot asks the user for the fields and criteria with which performing the query. The allowed operators for the filtering criteria are described by the chatbot. Finally, in the right fragment, the chatbot shows the results of the query, and presents the user with two buttons to download the results in a CSV plain text file or a Microsoft Excel spreadsheet.

5.1.2. Evaluation

We report a conducted a user study aimed to evaluate the chatbot according to the achievement of a number of public service values, as well as measuring distinct objective and subjective metrics. To design our study, we considered previous works that evaluate chatbots in e-government (Aoki, 2020; Petriv et al., 2019), and works that survey evaluation methodologies and metrics for conversational systems (Maroengsit et al., 2019; Peras, 2018; Ren, Castro, Acuña, & de Lara, 2019).

Regarding the evaluation method, the study was done in a controlled setting. A total of 12 people participated. There were 4 females and 8 males, ranging in age from 18 to 54 years old. All of them had used web search engines frequently, and only one had not used a chatbot. Participants had diverse levels of knowledge/expertise on spreadsheets –low (2), medium (8) and high (2)– and SQL –null (5), low (5) and high (2).

We aimed to assess our chatbot in comparison to the traditional method followed to consume open data. Participants were split into two groups for searching and exploring open data collections: one group (the control group) used the portal's search engine and Microsoft Excel spreadsheet application, and the other group (the experimental group) used our chatbot. They were requested to perform three tasks of increasing difficulty that, without the chatbot, entails processing operations on the spreadsheets to find the target information: 1) finding the public holidays in 2021; 2) finding the name and salary of the 10 councillors who earned the most money in 2020; and 3) finding the total money allocated in the budget 2020 for culture activities.

We considered both objective and subjective metrics. On the one hand, all actions performed by participants on the systems (i.e., portal and chatbot) were recorded in log files, to be able to afterward analyze system performance aspects, such as *effectiveness* and *efficiency*. On the other hand, after finishing a task, participants were requested to fill an intermediate task-oriented questionnaire to gather comments and opinions about the perceived degree of difficulty of the tasks, and the *utility* of the used tools.

In addition to these metrics, as a novel contribution of our work, we considered the theoretical framework recently proposed by Makasi et al. (2020), aimed to establish the public values a chatbot-mediated public service should have. In particular, we implemented the framework as a final questionnaire with 14 items, aimed to evaluate whether each considered data consumption method (portal+spreadsheets vs. chatbot) met the proposed public values.

The following are some results of our evaluation. Regarding *effectiveness*, the chatbot allowed participants to correctly complete the three tasks. Using the portal and spreadsheets, in contrast, only 66.7% of the task attempts were successfully completed, and 11.1% were not even completed. With respect to *efficiency*, using the portal, there was an increasing time for performing the three tasks (4.8, 7.2 and 9.9 min.), which is in accordance with their level of difficulty. This increment did not occur using the chatbot, with which the tasks were completed in similar times (3.4, 3.3 and 4.6 min.).

Regarding *usability*, the group using the portal, in general, responded that it was difficult or easy, while the group using the chatbot, in their majority, perceived that performing such tasks was easy or very easy.

Some of the results achieved for public value-based metrics were the following. The chatbot received more positive opinions than the portal in terms of *openness* and *accountability*. This, according to the participants' comments, was due to the explanations given by the chatbot in the whole data access process. This also applies to the *fairness* value, for which some participants expressed doubts about the capability of the portal search engine to retrieve the collections related to the user queries. All participants confirmed their *trust* on the veracity and *legitimacy* of the data used. By contrast, there were some concerns on the *privacy* aspect by participants who raise doubts about how browsing cookies were used.

5.2. A chatbot for exploring citizen participation content

As observed in the surveyed literature, chatbots have been recognized as powerful tools for e-participation applications, but so far have

[7] SQL stands for the Structured Query Language for relational databases.
[8] Apache Lucene, Java library for text indexing and search, https://lucene.apache.org
[9] Open Data portal of Madrid City Council, https://datos.madrid.es
[10] CSV stands for the Comma-Separated Values plain text storage format.

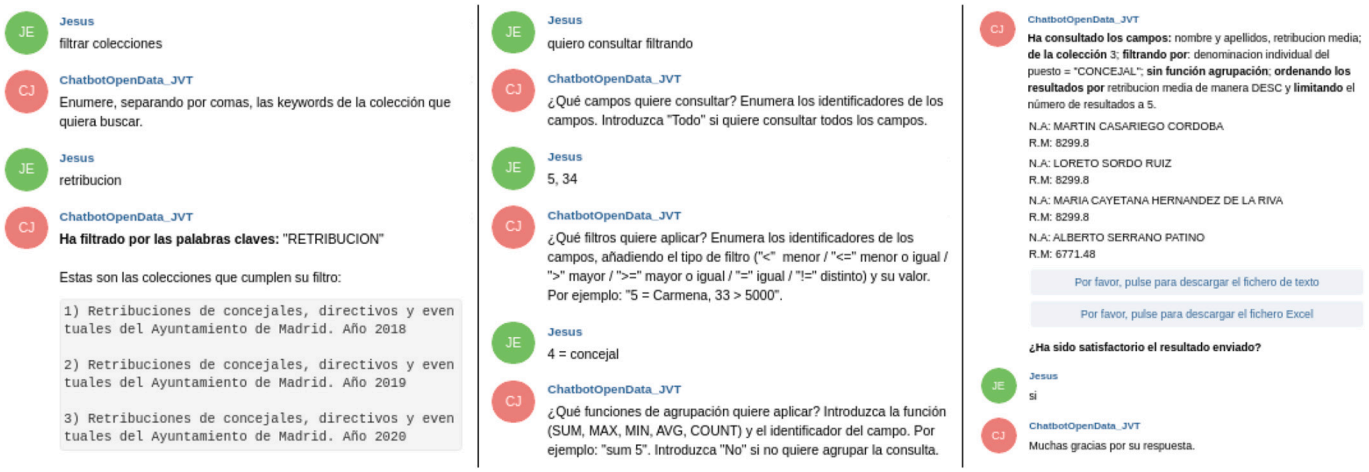

Fig. 3. Fragments of a conversation (in Spanish) with the open government data chatbot. In the first fragment, the user asks the chatbot for the list of data collections that match the keyword "retribución" (stands for "retribution" in English) and the chatbot responds with the list of 3 collections that satisfy the filter. In the second fragment, the user asks the chatbot for detailed information from the councilmen remuneration collection, and the chatbot, after performing an ad-hoc SQL query, responds with the requested data to the user.

been mainly proposed as discussion facilitators, and in general have been evaluated in terms of user involvement and engagement.

Motivated by this situation, we developed a chatbot to support the exploration of citizen-generated content in e-participation tools. As a novel contribution, our chatbot uses argument mining methods to extract and visualize argumentative information underlying the citizens' proposals and debates. This information is used to guide the users' navigation, and could be exploited in the discussion process as well.

To evaluate the chatbot, we considered a corpus from the 'Decide Madrid' e-participatory budgeting platform (Cantador, Cortés-Cediel, & Fernández, 2020). The dataset contained information about 21,744 citizen proposals, automatically classified into 30 categories and 325 topics, geolocated in 21 city districts, and annotated with controversy scores.

5.2.1. *Conversation intents*

The conversation flow handled by our chatbot is composed of eight intents. A couple of these intents allow listing the categories and topics

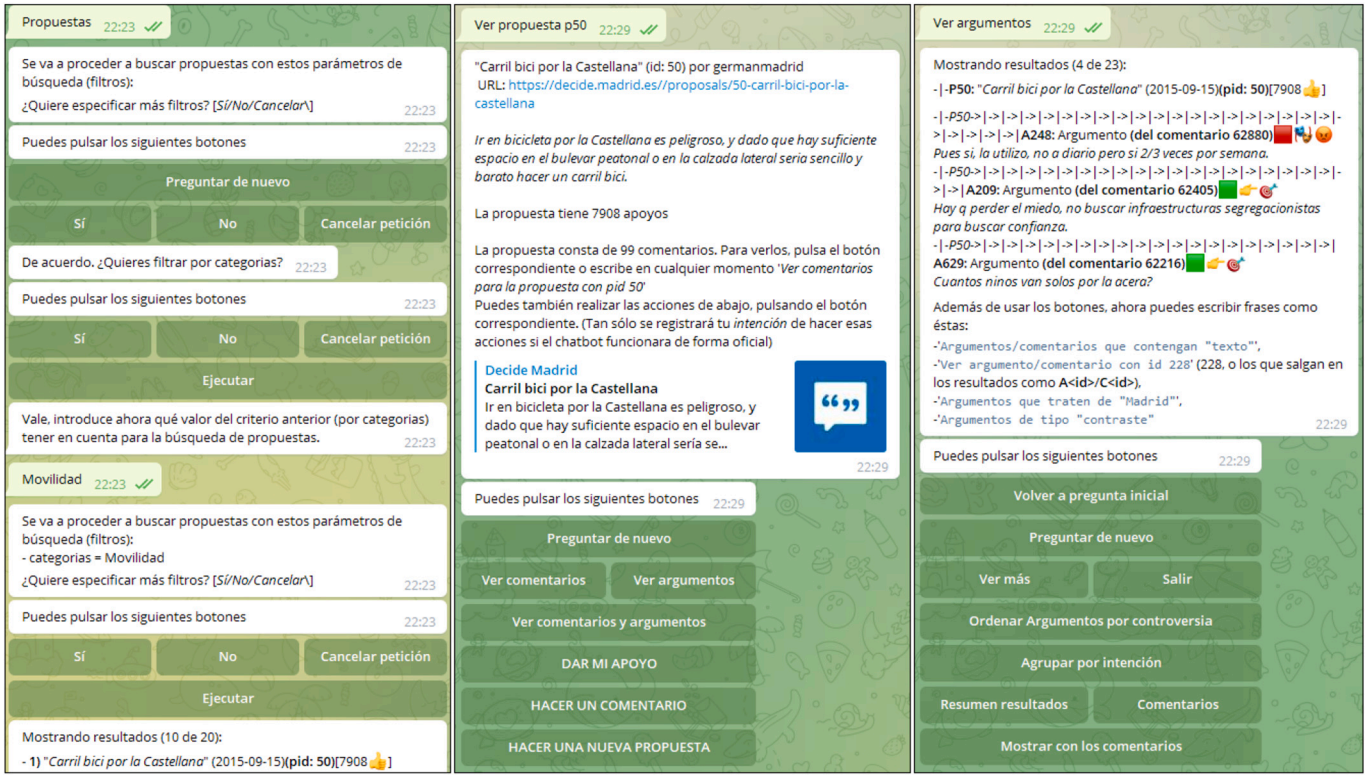

Fig. 4. Fragments of a conversation (in Spanish) with the citizen participation chatbot. In the first fragment, the user queries the chatbot for citizen proposals belonging to the category "Movilidad" (stands for "mobility" in English), and the chatbot shows the titles and ids of the 20 proposals that satisfy that filter. In the second fragment, the user requests information about a specific proposal (p50), and the chatbot returns the title, description and number of comments and supports of that proposal. In the third fragment, the user queries the chatbot for the arguments in favor or against proposal 50, and the chatbot displays them in a user-friendly format.

available in the platform. The user can ask for the citizen proposals of a given category or topic. Example of sentences that trigger this intent are 'Proposals of category…" and "Which are the proposals with topic…" Once a list of proposals is presented by the chatbot, a numerical identifier is shown for each proposal. With an identifier, the user can ask for the data associated to the corresponding proposal: title, summary and number of votes.

The chatbot provides several buttons that allow the user to access the comments and arguments of a proposal, and give several types of feedback: voting the proposal, making a comment, and creating a new proposal. This intent can also be triggered at any time when sentences like "Show comments of proposal with id…" are introduced. If a proposal search was recently executed, utterances like "Comments from last proposals" are recognized as well. The intent allows exploring iteratively all the comments of a given proposal or list of proposals. The intent about proposal arguments offers analogous functionalities to the comments intent, but applied to arguments extracted from proposal comments. In this case, an additional type of utterance is allowed, which is grouping arguments by topic.

Fig. 4 shows three screenshots with fragments of a conversation maintained in the chatbot. From left to right, they show a filtering process of proposals, details of a given proposal, and a set of categorized arguments existing in a proposal's comments. In the latter case, the intent, type and subtype of each argument are depicted through representative emojis.

5.2.2. Evaluation

We report a user study aimed at evaluating the developed chatbot and assessing the benefits of using argument-driven information exploration in e-participation with respect to a traditional topic keyword-based navigation.

A total of 32 people participated in the study: 22 male and 10 female of ages ranging 18–60 years old, and with different education levels. They had relatively low levels of knowledge/expertise on chatbots –null knowledge and/or expertise (10), low expertise (20), and medium expertise (2)– and on citizen participation –null (7), low (16) and medium (9).

Participants were randomly and uniformly split into two groups: a control group whose members only used the topic-driven (i.e., non argument-driven) browsing commands of our chatbot, and an experimental group whose members also used the chatbot argument-driven browsing commands.

Broadly speaking, we hypothesized that users of the experimental group would use the chatbot to a greater extent, and would make more positive opinions about the chatbot. To validate these hypotheses, we conducted both offline and online experimentation. With respect to the offline evaluation, all user interactions with the chatbot were recorded as time stamped logs in a database. After the one-week testing phase, the recorded logs were used to measure a variety of metrics related to the users' activity and engagement on the chatbot. Regarding the online evaluation, at the end of the testing phase, participants filled an opinion questionnaire aimed to measure the perceived system performance, citizen participation, and public value criteria, based on the framework proposed by Makasi et al. (2020).

The *engagement* results of the offline evaluation show that, although there was no significant difference in the average number of sessions per user between groups (2.8 in both cases), the sessions of the experimental group were longer than the sessions of the control group. Specifically, there was an increase of 45.6% on the average session duration, as well as an increase of 14.3% (from 56.8 to 64.9) on the average number of actions per user, and of 23.5% (from 1.7 to 2.1) on the average number of feedback provision actions per user.

Regarding the online evaluation, there were no significant differences between the control and experimental groups with respect to the perception of *ease of use* and *efficiency* of the chatbot. By contrast, important differences were obtained in the potential utility of the chatbot: *usability* for exploring the citizen-generated content, *usefulness* for finding out and understanding existing citizens' opinions, and *persuasiveness* for promoting citizen participation. For these evaluation criteria, the experimental group expressed higher scores: 4.3 vs. 4.8, 4.0 vs. 4.5, and 3.8 vs. 4.7, respectively.

A similar trend was observed on the perceived levels of public value-based metrics: *transparency* and *fairness*. In these cases, the argument-driven instantiation of the chatbot achieved the highest score differences with respect to the non-argumentative version: 4.0 vs. 4.8, and 3.9 vs. 4.8, respectively. Finally, *satisfaction* and *engagement* were equally and positively evaluated in the two versions of the chatbot. However, they got more moderate scores in comparison to other criteria.

5.3. Limitations of the conducted studies

We believe that the results obtained in the studies we conducted are promising and bring several insights on how chatbots that access open government data and citizen participation content can be evaluated. Moreover, we claim a contribution to the combination of objective and subjective metrics to evaluate different aspects of e-government chatbots. However, we recognize that our studies are preliminary, and should be improved and extended, especially with the participation of a large number of users, and possibly considering additional tasks to be performed in the studies, which would entail more significant and generalizable results.

6. Conclusions

In this paper, we have conducted a thorough, holistic survey of 33 e-government chatbots that have been proposed in the research literature and 25 chatbots that are currently being used by the Spanish public administration. With our review, we have identified and discussed major trends and challenges of both sets of chatbots (RQ1), as well as some research gaps between them (RQ2). Moreover, we have presented and evaluated a couple of chatbots for exploring open government data and citizen participation content, which represent significant advances in the state of the art (RQ3).

Among the reported open issues, we highlight the potential of developing novel G2G and G2B applications for chatbots operating with government and enterprise stakeholders, respectively; the incorporation of personalization in the public services offered by the chatbots; and the formalization of metrics and methodologies to evaluate the chatbots.

In addition to the identified research challenges and gaps, we here take the opportunity to comment on several ambitious, open issues that may raise a new generation of chatbots aimed to create public value. The first one is the **reuse and fast development** of chatbot implementations. Even using recent conversational agent technologies, implementing a chatbot from scratch is a difficult and time-consuming task. In this context, recent approaches to automatic generation of chatbots could be exploited (Ed-Douibi, Cánovas Izquierdo, Daniel, & Cabot, 2021; Pérez-Soler, Guerra, & De Lara, 2020), allowing researchers to focus on design and evaluation aspects related to the underlying public services and factors. This is important because recent research has put emphasis on the need of analyzing users experience to design chatbots to achieve higher public values (Song, 2022). Although prior work has indicated a total of 82 requirements related to usefulness, ease of use and presence to aid the design of these tools (Mafra et al., 2022), those specially driven to improving efficiency, effectiveness, sustainability, trust (Song, 2022), privacy and security (Brüggemeier & Lalone, 2022) are really relevant for public value creation.

Second, future research and development of e-government chatbots could take advantage of the unprecedented, impressive **advances in NLP** in general, and in conversational agents in particular. Recent research indicates that chatbots in public services should entail affording the socially vulnerable opportunities to participate in public affairs (Song, 2022), mainly to bridge the digital divide, which exists even with

the use of AI systems (Valle-Cruz, Alejandro Ruvalcaba-Gomez, Sandoval-Almazan, & Ignacio Criado, 2019). In this context, modern NLP models are really relevant for efficient chatbots to solve digital divide issues, as recently exemplified by the already well-known ChatGPT system (van Dis, Bollen, Zuidema, van Rooij, & Bockting, 2023), which has the ability to generate a wide range of detailed text responses for freely stated, open-domain questions and requests.

Third, future chatbots may get closer to being expert systems, incorporating or integrating **decision-making capabilities** (Watson, 2017). They could have and exploit knowledge to guide the user to the solution of a problem, proactively asking questions and leading the conversation flow. Related to this, chatbots have been envisioned to support collaborative work (Følstad et al., 2021), and recent research has indicated the need for AI-based feedback systems to lead citizens to more comprehensible argumentation on urban participation platforms (Borchers, Tavanapour, & Bittner, 2023), which could help increase citizen participation in new, more collaborative governance models (Rodríguez Bolívar, 2022). Hence, within the context of citizen participation, chatbots may assist in idea generation cooperative processes. Acting as a mediator or even collaborator of a group of (different) stakeholders, a chatbot may mediate between participants, search for similar problems or solutions to those discussed by the group, or complement given opinions and arguments with external data, to name a few potential functionalities.

Moreover, as in AI and other computer science disciplines (Shin & Park, 2019), for future e-government chatbots, addressing the so-called **FATE dimensions** (fairness, accountability, transparency and ethics) may represent a priority. These desirable features should not only be considered in the evaluation of chatbots, but also in their design and implementation. Indeed, prior research has highlighted a great number of risks with AI and chatbots use (Yang, Chen, Por, & Ku, 2023), mainly based on ethical and legitimacy challenges (e.g., moral dilemmas (Wirtz, Weyerer, & Geyer, 2019), unethical exploitation of data (Fatima, Desouza, & Dawson, 2020), AI discrimination (de Sousa, de Melo, Bermejo, Farias, & Gomes, 2019), and data sharing in healthcare (Sun & Medaglia, 2019), privacy and security issues (e.g., security threads (Edu et al., 2022), malicious inputs (Adamopoulou & Moussiades, 2020), user profiling, contextual attacks and data breaches (Ye & Li, 2020), trust (Sun & Medaglia, 2019), and unfairness in the delivery of public services (Chen, Ran, & Gao, 2019).

Prior research points out that some of these **risks** can be mitigated by both modeling algorithm literacy to increase user trust in chatbot applications (Shin, 2022) and using other emerging technologies jointly with the chatbot implementation (Yang et al., 2023). Therefore, future research should deepen into these insights designing joint emerging technologies deployment and investigating in greater detail conceptual links between literacy, trust and credibility for chatbot implementations with the aim at solving all the risks identified.

All the benefits and risks concerning chatbot use are very relevant for public managers because they entail uncertainty and challenges for the chatbot adoption (Sienkiewicz-Małyjurek, 2023). Indeed, public managers have been project managers of AI implementations in most of the cases (Tangi, van Noordt, & Rodriguez Müller, 2023) and, therefore, play a pivotal role in properly implementing AI in public administrations, leading them to work in many perspectives from two different approaches: piloting and implementing. They are driven by intrinsic motivation to perform more qualified activities, which makes them put themselves at the forefront of technological innovation (Maragno, Tangi, Gastaldi, & Benedetti, 2022).

Besides, the perception of AI adoption by **public managers** is crucial (Sienkiewicz-Małyjurek, 2023), since they must manage the cultural change that chatbots may introduce so as not to be perceived as a threat (Maragno et al., 2022). In this regard, the AI implementation in public administration brings many challenges, including legal, political, organizational and ethical issues (Sienkiewicz-Małyjurek, 2023; Tangi et al., 2023), such as insufficient AI regulations (Sienkiewicz-Małyjurek, 2023), the lack of strategic planning for AI implementation (Dwivedi et al., 2021), or difficulties in finding highly qualified staff (de Sousa et al., 2019). In addition, public managers must coordinate agents' activities, ensuring that all actors (humans and chatbots) have the necessary information to individually perform their tasks with the aim of reaching a successful final output (Maragno et al., 2022). If everything is well-managed, chatbots would have allowed public managers to relieve civil servants from repetitive tasks (Maragno et al., 2022). Hence, to achieve successful AI implementations in public administrations, public managers need to focus not only on external strategy-based planning, but also on internal capabilities building (Madan & Ashok, 2022).

Finally, having a holistic approach for developing chatbots in public administrations, as our research provides, can be considered as a fundamental prerequisite for understanding the underlying digital transformation. This holistic view, however, has some **limitations** that may lead to new research avenues. First, the exploratory nature of our work from the social sciences' perspective requires deeper analyses concerning the external validity of our results and the potential generalization of our conclusions. Second, our analysis was performed in a particular context: the Spanish public administration case. As noted by Dwivedi et al. (2023), generative AI implementation should be mapped to its context of use and application. Therefore, future research should deepen in other countries and/or specific sectors (health/medical sector, education, immigration, etc.) to gather more specific insights concerning the chatbot adoption. For example, recent research in Germany (Rude & Giesing, 2022) has confirmed that AI implementation could increase the wage and unemployment gaps of the migrant and native population. Future research should thus deepen this effect in other countries or in different types of applications. Also, AI offers the potential for security authorities to counter hybrid threats to preserve territorial integrity and protect the population (Androniceanu, 2023). In this regard, scholars have to put greater attention in designing better security in chatbots. According to all these aspects, our research has only pointed out some future avenues based on a holistic view of the topic, jointly using an analysis of literature review and experiences in chatbot deployments, and the creation and evaluation of two specific chatbots for information and participation.

CRediT authorship contribution statement

María E. Cortés-Cediel: Conceptualization, Formal analysis, Investigation, Writing – original draft. **Andrés Segura-Tinoco:** Conceptualization, Formal analysis, Writing – original draft. **Iván Cantador:** Conceptualization, Methodology, Investigation, Writing – original draft, Supervision, Visualization, Funding acquisition. **Manuel Pedro Rodríguez Bolívar:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation (PID2019-108965GB-I00) and the Regional Government of Andalusia (P20_00314 and B-SEJ-556-UGR20). The authors thank all people who participated in the reported studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.giq.2023.101877.

References

Abbas, N., Følstad, A., & Bjørkli, C. A. (2023). Chatbots as part of digital government service provision – A user perspective. In *Proceedings of the 3rd south American international industrial engineering and operations management conference* (pp. 66–82). Springer.

Acer, U. G., Van den Broeck, M., & Kawsar, F. (2019a). The city as a personal assistant. In *Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers* (pp. 1102–1106). ACM.

Acer, U. G., Van den Broeck, M., & Kawsar, F. (2019b). On hyper-local conversational agents in urban settings. In *Proceedings of the 5th ACM workshop on Mobile Systems for Computational Social Science* (pp. 1–4).

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications, 2*, Article 100006.

Adnan, S. M., Hamdan, A., & Alareeni, B. (2021). Artificial intelligence for public sector: Chatbots as a customer service representative. In *Proceedings of the 1st international conference on business and technology* (pp. 164–173). Springer.

Afonasova, M. A., Panfilova, E. E., Galichkina, M. A., & Ślusarczyk, B. (2019). Digitalization in economy and innovation: The effect on social and economic processes. *Polish Journal of Management Studies, 19*, 22–32.

Ahmady, S. E., & Uchida, O. (2020). Telegram-based chatbot application for foreign people in Japan to share disaster-related information in real-time. In *Proceedings of the 5th international conference on computer and communication systems* (pp. 177–181). IEEE.

Akkaya, C., & Krcmar, H. (2019). Potential use of digital assistants by governments for citizen services: The case of Germany. In *Proceedings of the 20th annual international conference on digital government research* (pp. 81–90). ACM.

Albert, A., & Passmore, E. (2008). *Public value and participation: A literature review for the Scottish government*. Scottish Government.

Alshehhi, K., Cheaitou, A., & Rashid, H. (2022). Adoption frameworks for artificial intelligence in the public sector: A systematic review of literature. In *Proceedings of the 6th international workshop on Chatbot research, Applications and Design* (pp. 919–929). IEOM Society International.

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration, 97*, 296–310.

Androniceanu, A. (2023). The new trends of digital transformation and artificial intelligence in public administration. *Administratie si Management Public*, 147–155.

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly, 36*, 358–367.

Anelli, V. W., Di Noia, T., Di Sciascio, E., & Ragone, A. (2019). Anna: A virtual assistant to interact with Puglia digital library (discussion paper). In *Proceedings of the 27th Italian symposium on advanced database systems* (pp. 30:1–30:8). CEUR Workshops.

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly, 37*, Article 101490.

Arifin, A. A., & Lennerfors, T. T. (2021). Ethical aspects of voice assistants: A critical discourse analysis of Indonesian media texts. *Journal of Information, Communication and Ethics in Society, 20*, 18–36.

Atreja, S., Aggarwal, P., Mohapatra, P., Dumrewal, A., Basu, A., & Dasgupta, G. B. (2018). Citicafe: An interactive interface for citizen engagement. In *Proceedings of the 23rd international conference on intelligent user interfaces* (pp. 617–628). ACM.

Augello, A., Pilato, G., & Gaglio, S. (2009). An intelligent agent to support city policies decisions. In *Proceedings of the 19th Italian workshop of the Italian Society for Neural Networks* (pp. 163–168). IOS Press.

Augello, A., Pilato, G., & Gaglio, S. (2011). Intelligent advisor agents in distributed environments. In *Information retrieval and Mining in Distributed Environments* (pp. 109–124). Springer.

Ávila, C. V. S., Calixto, A. B., Rolim, T. V., Franco, W., Venceslau, A. D. P., Vidal, V. M. P., … De Moura, F. F. (2019). MediBot: An ontology based chatbot for Portuguese speakers drug's users. In *Proceedings of the 21st international conference on Enterprise information systems* (pp. 25–36). SciTePress.

Bagnasco, C., Cappelli, A., & Magnini, B. (2000). A dialogue environment for accessing public administration data: The TAMIC-P system. In *Proceedings of the 14th European conference on artificial intelligence* (pp. 686–690). IOS Press.

Baldauf, M., & Zimmermann, H. D. (2020). Towards conversational e-government: An experts' perspective on requirements and opportunities of voice-based citizen services. In *Proceedings of the 7th international conference HCI in business, government and organizations* (pp. 3–14). Springer.

Baldauf, M., Zimmermann, H. D., & Pedron, C. (2021). Exploring citizens' attitudes towards voice-based government services in Switzerland. In *Proceedings of the 23rd international conference on human-computer interaction, part III* (pp. 229–238). Springer.

Ballester, O. (2021). An artificial intelligence definition and classification framework for public sector applications. In *Proceedings of the 22nd annual international conference on digital government research* (pp. 67–75).

Bellini, V., Biancofiore, G. M., Di Noia, T., Di Sciascio, E., Narducci, F., & Pomo, C. (2020). Guapp: A conversational agent for job recommendation for the italian public administration. In *Proceedings of the 2020 IEEE conference on evolving and adaptive intelligent systems* (pp. 1–7). IEEE.

Benington, J., & Moore, M. (2010). *Public value: Theory and practice*. Bloomsbury Publishing.

Boden, C., Fischer, J., Herbig, K., & Spierling, U. (2006). CitizenTalk: Application of chatbot infotainment to e-democracy. In *Proceedings of the 3rd international conference on Technologies for Interactive Digital Storytelling and Entertainment* (pp. 370–381). Springer.

Borchers, M., Tavanapour, N., & Bittner, E. (2023). Exploring AI supported citizen argumentation on urban participation platforms. In *Proceedings of the 56th Hawaii international conference on system sciences* (pp. 1643–1652).

Brüggemeier, B., & Lalone, P. (2022). Perceptions and reactions to conversational privacy initiated by a conversational user interface. *Computer Speech & Language, 71*, Article 101269.

Cantador, I., Cortés-Cediel, M. E., & Fernández, M. (2020). Exploiting open data to analyze discussion and controversy in online citizen participation. *Information Processing & Management, 57*, Article 102301.

Cantador, I., Viejo-Tardío, J., Cortés-Cediel, M. E., & Rodríguez Bolívar, M. P. (2021). A chatbot for searching and exploring open data: Implementation and evaluation in e-government. In *Proceedings of the 22nd annual international conference on digital government research* (pp. 168–179). ACM.

Chen, T., Ran, L., & Gao, X. (2019). AI innovation for advancing public service: The case of China's first administrative approval bureau. In *Proceedings of the 20th annual international conference on digital government research* (pp. 100–108).

Chohan, S. R., Hu, G., Khan, A. U., Pasha, A. T., & Sheikh, M. A. (2021). Design and behavior science in government-to-citizens cognitive-communication: A study towards an inclusive framework. *Transforming Government: People, Process and Policy, 15*, 532–549.

Dahl, A., & Soss, J. (2014). Neoliberalism for the common good? Public value governance and the downsizing of democracy. *Public Administration Review, 74*, 496–504.

De Lacerda, A. R. T., & Aguiar, C. S. R. (2019). FLOSS FAQ chatbot project reuse: How to allow nonexperts to develop a chatbot. In *Proceedings of the 15th international symposium on open collaboration* (pp. 1–8). ACM.

Dhungel, A. K., Wessel, D., Zoubir, M., & Heine, M. (2021). Too bureaucratic to flexibly learn about AI? The human-centered development of a MOOC on artificial intelligence in and for public administration. In *Proceedings of mensch und computer 2021* (pp. 563–567). ACM.

van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*, 224–226.

Drydakis, N. (2021). Mobile applications aiming to facilitate immigrants' societal integration and overall level of integration, health and mental health. Does artificial intelligence enhance outcomes? *Computers in Human Behavior, 117*, Article 106661.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., , … Eirug, A., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, Article 101994.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., , … Ahuja, M., et al. (2023). "So what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management, 71*, Article 102642.

Ed-Douibi, H., Cánovas Izquierdo, J. L., Daniel, G., & Cabot, J. (2021). A model-based chatbot generation approach to converse with open data sources. In *Proceedings of the 21st international conference on web engineering* (pp. 440–455). Springer.

Edu, J., Mulligan, C., Pierazzi, F., Polakis, J., Suarez-Tangil, G., & Such, J. (2022). Exploring the security and privacy risks of chatbots in messaging services. In *Proceedings of the 22nd ACM internet measurement conference* (pp. 581–588).

Fatima, S., Desouza, K. C., & Dawson, G. S. (2020). National strategic artificial intelligence plans: A multi-dimensional analysis. *Economic Analysis and Policy, 67*, 178–194.

Federici, S., Mele, M. L., Bracalenti, M., De Filippis, M. L., Lanzilotti, R., Desolda, G., , … Amendola, M., et al. (2021). A chatbot solution for eGLU-box pro: The usability evaluation platform for Italian public administrations. In *Proceedings of the 23rd international conference human-computer interaction, part I* (pp. 268–279). Springer.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender bias in chatbot design. In *Chatbot research and design: Third international workshop, CONVERSATIONS 2019, Amsterdam, the Netherlands, November 19–20, 2019, revised selected papers 3* (pp. 79–93). Springer.

Firdaus, V., Saputra, P., & Suprianto, D. (2020). Intelligence chatbot for Indonesian law on electronic information and transaction. In *Proceedings of the 3rd international conference on innovation in engineering and vocational education* (p. 022089). IOP Publishing.

Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., et al. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing, 103*, 2915–2942.

Ford, H., & Hutchinson, J. (2019). Newsbots that mediate journalist and audience relationships. *Digital Journalism, 7*, 1013–1031.

Gatius, M., González, M., & Comelles, E. (2007). An information state-based dialogue manager for making voice web smarter. In *Proceedings of the 16th international conference on world wide web* (pp. 1315–1316). ACM.

Gatius, M., González, M., Militello, S., & Hernández, P. (2006). Integrating semantic web and language technologies to improve the online public administrations services. In *Proceedings of the 15th international conference on world wide web* (pp. 847–848). ACM.

Gerontas, A., Zeginis, D., Promikyridis, R., Androš, M., Tambouris, E., Cipan, V., & Tarabanis, K. (2022). Enhancing core public service vocabulary to enable public service personalization. *Information, 13*, 225.

Griol, D., & García-Jiménez, M. (2012). Development of interactive virtual voice portals to provide municipal information. In *Proceedings of the 9th international conference on distributed computing and artificial intelligence* (pp. 161–172). Springer.

Guy de Andrade, G., Ramos Sousa Silva, G., Molina Duarte Júnior, F. C., Almeida Santos, G., Lopes de Mendonça, F. L., & Timóteo de Sousa Júnior, R. (2020). EvaTalk:

A chatbot system for the Brazilian government virtual school. In *Proceedings of the 22nd international conference on Enterprise information systems* (pp. 556–562). SciTePress.
Haqbeen, J., Sahab, S., Ito, T., & Rizzi, P. (2021). Using decision support system to enable crowd identify neighborhood issues and its solutions for policy makers: An online experiment at Kabul municipal level. *Sustainability, 13*, 5453.
Hasan, I., Rizvi, S., Jain, S., & Huria, S. (2021). The AI enabled chatbot framework for intelligent citizen-government interaction for delivery of services. In *Proceedings of the 8th international conference on computing for sustainable global development* (pp. 601–606). IEEE.
Kalampokis, E., Karacapilidis, N., Tsakalidis, D., & Tarabanis, K. (2023). *Understanding the use of emerging technologies in the public sector: A review of horizon 2020 projects*. Digital Government: Research and Practice.
Kamoen, N., & Liebrecht, C. (2022). I need a CAVAA: How conversational agent voting advice applications (cavaas) affect users' political knowledge and tool experience. *Frontiers in Artificial Intelligence, 5*, Article 835505.
Keyner, S., Savenkov, V., & Vakulenko, S. (2019). Open data chatbot. In *Proceedings of the 16th European semantic web conference* (pp. 111–115). Springer.
Kuberkar, S., & Singhal, T. K. (2020). Factors influencing adoption intention of AI powered chatbot for public transport services within a smart city. *International Journal of Emerging Technologies in Learning, 11*, 948–958.
Kucherbaev, P., Psyllidis, A., & Bozzon, A. (2017). Chatbots as conversational recommender systems in urban contexts. In *Proceedings of the 1st international workshop on recommender systems for citizens* (pp. 1–2). ACM.
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine, 151*. W–65.
Lockett, J., Wijesinghe, S., Phillips, J., Gross, I., Schoenfeld, M., Hiranpat, W. T., … Hu, Q. (2019). Intelligent voice agent and service (iVAS) for interactive and multimodal question and answers. In *Proceedings of the 13th international conference on flexible query answering systems* (pp. 396–402). Springer.
Lommatzsch, A. (2018). A next generation chatbot-framework for the public administration. In *Proceedings of the 18th international conference on innovations for community services* (pp. 127–141). Springer.
Lommatzsch, A., & Katins, J. (2019). An information retrieval-based approach for building intuitive chatbots for large knowledge bases. In *Proceedings of the 2019 conference Lernen, Wissen, Daten, Analysen* (pp. 343–352). CEUR Workshops.
Madan, R., & Ashok, M. (2022). Ai adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly, 101774*.
Mafra, M., Nunes, K., Castro, A., Lopes, A., Oran, A. C., Braz Junior, G., , … Rocha, S., et al. (2022). Defining requirements for the development of useful and usable chatbots: An analysis of quality attributes from academy and industry. In *Proceedings of the 22nd international conference on human-computer interaction* (pp. 479–493). Springer.
Mahapatra, R., Sharma, N., Trivedi, A., & Aman, C. (2012). Adding interactive interface to e-government systems using AIML based chatterbots. In *Proceedings of the 6th international conference on software engineering* (pp. 1–6). IEEE.
Makasi, T., Nili, A., Desouza, K., & Tate, M. (2020). Chatbot-mediated public service delivery. *First Monday, 12*.
Makasi, T., Nili, A., Desouza, K., & Tate, M. (2022). Public service values and chatbots in the public sector: Reconciling designer efforts and user expectations. In *Proceedings of the 55th Hawaii international conference on system sciences* (pp. 2334–2343). AIS.
Makasi, T., Nili, A., Desouza, K. C., & Tate, M. (2021). A typology of chatbots in public service delivery. *IEEE Software, 39*, 58–66.
Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2021). The spread of artificial intelligence in the public sector: A worldwide overview. In *Proceedings of the 14th international conference on theory and practice of electronic governance* (pp. 1–9). ACM.
Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2022). AI as an organizational agent to nurture: Effectively introducing chatbots in public entities. *Public Management Review*, 1–31.
Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. In *2019 7th international conference on information and education technology* (pp. 111–119). ACM.
Martin, A., Allagha, M., & Misnikov, Y. (2021). Emyconnect: Building a smart public service for young mobile Europeans. In *Proceedings of the 2021 Conference on advances in the human side of service engineering* (pp. 37–44). Springer.
Massai, L., Nesi, P., & Pantaleo, G. (2019). PAVAL: A location-aware virtual personal assistant for retrieving geolocated points of interest and location-based services. *Engineering Applications of Artificial Intelligence, 77*, 70–85.
Meng, A., & Khelladi, Y. (2017). Comparative use of web form, SMS, and chatbot in social election monitoring of the Dominican 2016 general election. In *Proceedings of the 9th international conference on information and communication technologies and development* (pp. 1–5).
Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. *npj Digital Medicine, 3*, 65:1–65:4.
Morse, R. S. (2010). Integrative public leadership: Catalyzing collaboration to create public value. *The Leadership Quarterly, 21*, 231–245.
Neumaier, S., Savenkov, V., & Vakulenko, S. (2017). Talking open data. In *ESWC 2017 Satellite Events* (pp. 132–136). Springer.
Nirala, K. K., Singh, N. K., & Purani, V. S. (2022). A survey on providing customer and public administration based services using AI: Chatbot. *Multimedia Tools and Applications, 81*, 22215–22246.
Nowakowska-Grunt, J., Dziadkiewicz, M., Olejniczak-Szuster, K., & Starostka-Patyk, M. (2021). Quality of service in local government units and digital exclusion of elderly people-example from implementing the avatar project. Polish. *Journal of Management Studies, 23*.
Pal, D., Arpnikanondt, C., Razzaque, M. A., & Funilkul, S. (2020). To trust or not-trust: Privacy issues with voice assistants. *IT Professional, 22*, 46–53.
Panagiotopoulos, P., Klievink, B., & Cordella, A. (2019). *Public value creation in digital government*.
Peras, D. (2018). Chatbot evaluation metrics. In *36th international scientific conference on economic and social development, Varazdin development and entrepreneurship agency* (pp. 89–97).
Pérez-Soler, S., Guerra, E., & De Lara, J. (2020). Model-driven chatbot development. In *Proceedings of the 39th international conference on conceptual modeling* (pp. 207–222). Springer.
Petriv, Y., Erlenheim, R., Tsap, V., Pappel, I., & Draheim, D. (2019). Designing effective chatbot solutions for the public sector: A case study from Ukraine. In *Proceedings of the 6th international conference on electronic governance and open Society* (pp. 320–335). Springer.
Piccolo, L. S. G., Roberts, S., Iosif, A., & Alani, H. (2018). Designing chatbots for crises: A case study contrasting potential and reality. In *Proceedings of the 32nd international BCS human computer interaction conference* (pp. 56:1–56:10). ACM.
Porreca, S., Leotta, F., Mecella, M., & Catarci, T. (2017). Chatbots as a novel access method for government open data. In *Proceedings of the 25th Italian symposium on advanced database systems* (pp. 19:1–19:8). CEUR Workshops.
Porreca, S., Leotta, F., Mecella, M., Vassos, S., & Catarci, T. (2018). Accessing government open data through chatbots. In *Proceedings of the ICWE 2017 international workshops* (pp. 156–165). Springer.
Portela, M. (2021). Interfacing participation in citizen science projects with conversational agents. *Human Computation, 8*, 33–53.
Ramires Hernández, P., Valle-Cruz, D., & Méndez Mendoza, R. V. (2023). Review on the application of artificial intelligence-based chatbots in public administration. In *Handbook of research on applied artificial intelligence and robotics for government processes* (pp. 133–155). IGI Global.
Rauschenberger, M., Schrepp, M., Cota, M. P., Olschner, S., & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (UEQ). Example: Spanish language version. *International Journal of Artificial Intelligence and Interactive Multimedia, 2*, 49–45.
Ren, R., Castro, J. W., Acuña, S. T., & de Lara, J. (2019). Evaluation techniques for chatbot usability: A systematic mapping study. *Intl Journal of Software Engineering and Knowledge Engineering, 29*, 1673–1702.
Rodríguez Bolívar, M. P. (2017). Policy makers' perceptions on the transformational effect of web 2.0 technologies on public services delivery. *Electronic Commerce Research, 17*, 227–254.
Rodríguez Bolívar, M. P. (2022). Learning from the COVID-19 pandemic in governing smart cities. *Journal of Smart Cities and Society*, 1–15.
Ronzhin, S., Folmer, E., Maria, P., Brattinga, M., Beek, W., Lemmens, R., & van't Veer, R. (2019). Kadaster knowledge graph: Beyond the fifth star of open data. *Information, 10*, 310.
Rozenes, S., & Cohen, Y. (2022). Artificial intelligence synergetic opportunities in services: Conversational systems perspective. *Applied Sciences, 12*, 8363.
Rude, B., & Giesing, Y. (2022). Technological change and immigration - a race for talent or of displaced workers. In *Technical Report*. Verein für Socialpolitik / German Economic Associatio.
Sánchez-Nielsen, E., Morales, A., Mendo, O., & Chávez-Gutiérrez, F. (2021). SuDaMa: Sustainable open government data management framework for long-term publishing and consumption. *IEEE Access, 9*, 151841–151863.
Sandoval-Almazán, R., & Gutiérrez-Alonso, M. A. (2009). Virtual assistants for e-government interaction. In *Social and political implications of data mining: Knowledge management in e-government* (pp. 255–266). IGI Global.
Sandoval-Almazán, R., & Millán-Vargas, A. O. (2023). Artificial intelligence in innovation labs: Map of cases for the public sector. In *Handbook of research on applied artificial intelligence and robotics for government processes* (pp. 115–132). IGI Global.
Segura-Tinoco, A., Holgado-Sánchez, A., Cantador, I., Cortés-Cediel, M. E., & Rodríguez Bolívar, M. P. (2022). A conversational agent for argument-driven e-participation. In *Proceedings of the 23rd annual international conference on digital government research* (pp. 191–205). ACM.
Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & Society*, 1–16.
Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284.
Sienkiewicz-Małyjurek, K. (2023). Whether AI adoption challenges matter for public managers? The case of polish cities. *Government Information Quarterly, 101828*.
Simonsen, L., Steinstø, T., Verne, G., & Bratteteig, T. (2020). "I'm disabled and married to a foreign single mother". Public service chatbot's advice on citizens' complex lives. In *Proceedings of the 12th IFIP WG 8.5 international conference on electronic participation* (pp. 133–146). Springer.
Song, J. (2022). The implications of providing voice-based chatbots in public service for digital inclusion and public communication. *OMNES: The Journal of Multicultural Society, 12*, 26–68.
de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly, 36*, Article 101392.
Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36*, 368–383.

Szostak, M. (2022). Peculiarities of art management in a digital context - a case study of Poland. *Entrepreneurship and Sustainability Issues, 9*, 10–44.
Tangi, L., van Noordt, C., & Rodriguez Müller, A. P. (2023). The challenges of AI implementation in the public sector. In *Proceedings of the 24th annual international conference on digital government research* (pp. 414–422). ACM.
Tanoue, Y., Nomura, S., Yoneoka, D., Kawashima, T., Eguchi, A., Shi, S., Harada, N., & Miyata, H. (2020). Mental health of family, friends, and co-workers of COVID-19 patients in Japan. *Psychiatry Research, 291*, Article 113067.
Tavanapour, N., Poser, M., & Bittner, E. A. C. (2019). Supporting the idea generation process in citizen participation-toward an interactive system with a conversational agent as facilitator. In *Proceedings of the 27th European conference on information systems* (pp. 70:1–70:17). AIS.
Thai, H. D., & Huh, J. H. (2021). Building an operational solution assistant system for foreign SMEs in ROK. *Applied Sciences, 11*, 4510.
Tisland, I., Sodefjed, M. L., Vassilakopoulou, P., & Pappas, I. O. (2022). The role of quality, trust, and empowerment in explaining satisfaction and use of chatbots in e-government. In *Proceedings of the 21st IFIP WG 6.11 conference on e-business, e-services and e-Society* (pp. 279–291). Springer.
Väänänen, K., Hiltunen, A., Varsaluoma, J., & Pietilä, I. (2020). Civicbots – Chatbots for supporting youth in societal participation. In *Proceedings of the 3rd international workshop on Chatbot research and design* (pp. 143–157). Springer.
Valle-Cruz, D., Alejandro Ruvalcaba-Gomez, E., Sandoval-Almazan, R., & Ignacio Criado, J. (2019). A review of artificial intelligence in government and its potential from a public policy perspective. In *Proceedings of the 20th annual international conference on digital government research* (pp. 91–99). ACM.
Valtolina, S., Barricelli, B. R., Gaetano, S. D., & Diliberto, P. (2018). Chatbots and conversational interfaces: Three domains of use. In *Proceedings of the 5th international workshop on cultures of participation in the digital age* (pp. 62–70). CEUR Workshops.
Valverde, M. S. D. R., et al. (2019). Chatbot in the online provision of government services. In *Proceedings of the 19th Portuguese association of information systems conference* (pp. 151–160). AIS.
Van Noordt, C., & Misuraca, G. (2019). New wine in old bottles: Chatbots in government – Exploring the transformative impact of chatbots in public service delivery. In *Proceedings of the 11th IFIP WG 8.5 international conference on electronic participation* (pp. 49–59). Springer.
Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Government Information Quarterly, 39*, Article 101714.
Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, O. I. (2022). Developing human/AI interactions for chat-based customer services: Lessons learned from the Norwegian government. *European Journal of Information Systems*, 1–13.
Vermaas, P., Kroes, P., Van de Poel, I., Franssen, M., & Houkes, W. (2022). *A philosophy of technology: From technical artefacts to sociotechnical systems*. Springer.
Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review, 80*, 946–961.
Watson, H. J. (2017). Preparing for the cognitive generation of decision support. *MIS Quarterly Executive, 16*.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration, 42*, 596–615.
Wright, J. (2021). The Alexafication of adult social care: Virtual assistants and the changing role of local government in England. *International Journal of Environmental Research and Public Health, 18*, 812.
Yang, J., Chen, Y. L., Por, L. Y., & Ku, C. S. (2023). A systematic literature review of information security in chatbots. *Applied Sciences, 13*, 6355.
Ye, W., & Li, Q. (2020). Chatbot security and privacy in the age of personal assistants. In *Proceedings of the 2020 IEEE/ACM symposium on edge computing* (pp. 388–393). IEEE.
Zogaj, A., Mähner, P. M., Yang, L., & Tscheulin, D. K. (2023). It'sa match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research, 155*, Article 113412.

María E. Cortés-Cediel is Associate Lecturer at the Faculty of Political Science and Sociology of Universidad Complutense de Madrid. She holds a BSc in Political Science and a BSc in Law from Universidad Autónoma de Madrid, and a PhD in Political Science, Public Administration and International Relations from Universidad Complutense de Madrid. Her research lines focus on citizen participation through technology, with special interest in the study of governance scenarios in smart cities. In the last five years, she has coauthored over 15 publications in international e-government journals and conferences, such as dg.o, IFIP EGOV-CeDEM-ePart, and HICSS.

Andrés Segura-Tinoco holds a BSc in Computer Science from Universidad de Oriente, Venezuela, and a MSc in Information Engineering from Universidad de Los Andes, Colombia. He has more than 14 years of experience working on data analysis projects, and creating intelligent and high-impact software solutions in global teams in USA, Mexico, Spain, and Malaysia., He is currently pursuing a PhD in Computer Science at Universidad Autónoma de Madrid. His research interests focus on recommender systems, natural language processing, and e-government.

Iván Cantador is Associate Professor of Computer Science at Universidad Autónoma de Madrid. He was a visiting researcher at The Open University, University of Southampton, University of Glasgow, and Free University of Bolzano. His research mainly lies in the areas of artificial intelligence, information retrieval, and recommender systems. He has investigated a wide range of issues related to user modeling, semantic representation of knowledge, and the processing and mining of user-generated content. He is member of the editorial board of the User Modeling and User-Adapted Interaction journal, and distinguished reviewer of the ACM Transactions of Recommender Systems journal. Dr. Cantador has co-authored more than 80 publications in international journals and conferences (e.g., SIGIR, ECIR, RecSys, and UMAP), where he has also collaborated as a member of numerous organizing and program committees.

Manuel Pedro Rodríguez Bolívar is Full Professor at the University of Granada. His areas of research are mainly related to information and technology in government, and, among others, include electronic and open government, public sector innovation, smart cities, and public policy evaluation. He has coauthored over 110 articles in national and international journals, of which more than 70 are published in journals indexed in SSCI/JCR, ABDC Journal Ratings and Journal Quality List. He is a member of the Editorial Board of 12 international journals (3 of them JCR in Q1 of Information and Library Science) and is part of the team of external evaluators of more than 60 journals, of which 47 are indexed in SSCI/JCR, and of 6 international conferences. He has been also the author of 56 book chapters, 49 of them published in prestigious international publishers such as Kluwer Academic Publishers, Springer, Routledge, Palgrave, Taylor and Francis and IGI Global, and is author of full-length books published by the Ministry of Economy and Finance in Spain. In addition, he is Editor in Chief of Sustainability (section sustainable management), IJPADA and series Editor of Public Administration and Information Technology book series in Springer. Finally, he has been collaborating with United Nations and he has co-chaired different tracks in international conferences such as ICEGOV, IFIP EGOV-CeDEM-ePart, HICSS and other conferences during the last years.