

A Conversational Agent for Argument-driven E-participation

ANDRÉS SEGURA-TINOCO, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain

ANDRÉS HOLGADO-SÁNCHEZ, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain

IVÁN CANTADOR, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain

MARÍA E. CORTÉS-CEDIEL, Facultad de Ciencias Políticas y Sociología, Universidad Complutense de Madrid, Spain

MANUEL PEDRO RODRÍGUEZ-BOLÍVAR, Facultad de Ciencias Económicas y Empresariales, Universidad de Granada, Spain

The majority of current e-participation tools are based on online web forums where citizens make proposals and provide comments and opinions, forming large conversation threads. Motivated by the huge popularity of instant messaging applications and the impressive, recent advances in natural language processing and artificial intelligence, in this paper we propose and investigate the use of conversational agents or chatbots as a new form of citizen-to-government communication. Specifically, we present and evaluate a novel chatbot that assists a user on the overwhelming task of exploring the citizen-generated content of Decide Madrid, a forum-based e-participatory budgeting platform. Among other things, the proposed chatbot is capable of automatically extracting, categorizing and summarizing the arguments underlying the citizen proposals and debates in the platform. Through a user study, we show promising results about the potential benefits of the chatbot in terms of several citizen participation, decision making and public value criteria.

CCS Concepts: • **Applied computing** → **E-government**; • **Computing methodologies** → **Discourse, dialogue and pragmatics**; • **Human-centered computing** → **Natural language interfaces**; **User studies**.

Additional Key Words and Phrases: e-government, participatory budgeting, chatbots, natural language processing, argument mining

ACM Reference Format:

Andrés Segura-Tinoco, Andrés Holgado-Sánchez, Iván Cantador, María E. Cortés-Cediel, and Manuel Pedro Rodríguez-Bolívar. 2022. A Conversational Agent for Argument-driven E-participation. In *dg.o '22: The 23rd Annual International Conference on Digital Government Research, June 15–17, 2022, Seoul, South Korea*. ACM, New York, NY, USA, 22 pages. <https://doi.org/XX.XXXX/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

Open Government represents a paradigm of public management that aims to facilitate the cooperation of citizens in the design and innovation of public services, and to strengthen the transparency and accountability of the public administration [19]. It has been proposed as a tool for collaborative governance that, from a perspective that empowers the citizens, has progressively promoted the use of technology as a means of interaction [16]. In fact, in recent years, the expansion of social media, the appearance of disruptive technologies –e.g., augmented and virtual reality, blockchain, and conversational agents–, and the consolidation of significant advances in the massive use of (open) data and artificial intelligence are supporting even more this type of governance [10].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

Particularly, citizen participation has been strongly influenced by new information and communication technologies, transforming the online participation and the democratization of the internet into major pillars of Open Government initiatives [23]. In this context, e-participation –understood as the computer-assisted support to citizen collaboration– has originated novel consultation and deliberation processes conducted through modern technologies, particularly web technologies.

The research literature on e-participation has mainly focused on the social consequences of using online platforms, rather than on the impact the platforms features have on the participation mechanisms and outcomes [17]. Hence, most research has been conducted with platforms based on traditional web forums, which have been the dominant type of tool used in e-participation. In the last years, nonetheless, attention has shifted to social media, especially social networks –e.g., Facebook and Twitter– [9, 30], and recently to instant messaging tools, such as Telegram and WhatsApp [4, 13].

Conventional web forums promote interaction over the production and reuse of collective knowledge. They offer smooth, large-scale interaction, but in general provide very limited or no functionalities for content organization, filtering and analysis [17]. This makes the moderation and processing of debate results (e.g., to create opinion summary reports) challenging and sometimes overwhelming. Furthermore, the forum content is structured by *time*, so it may be dispersed and redundant. Without additional mechanisms, such as reputation (rating) functionalities, it might be very difficult and time-consuming for users to find high quality contributions. Citizens would like to be able to process and understand the debates on certain topics in order to have global views of existing problems and proposed solutions, but forums do not usually have functionalities to support critical thinking and argumentation. As a consequence, contributions are often not guided by evidences, and thus limit the creation of informed and constructive debates [24].

Computer-supported argumentation visualization (CSAV) tools have been proposed to overcome the above shortcomings by providing structured deliberation support for e-participation [6, 7, 12, 17, 22, 25]. Specifically, these tools guide the deliberation around underlying arguments. They organize citizen-generated content by *topic* rather than by *time*, and represent a debate as a graph or network consisting of alternative positions on issues, as well as arguments in favor and against discussed ideas. Debates are then summarized as interactive maps displaying positions, arguments and links connecting them.

CSAV [18] thus favors knowledge representation over conversational interaction. It enables critical thinking and evidence-based reasoning, since it forces users to make explicit the rationale behind their claims. Besides, it fosters the deliberation on controversial issues as users can represent their points of view in coherent structures. However, CSAV also entails some disadvantages [17]. First, it may require users to undergo an intensive training to become proficient with the considered argument and debate models, and may involve higher coordination in terms of moderation and supervision of the argumentative maps. Second, it limits the social interaction and consequently may decrease the users' engagement. Third, large argumentation maps may hinder the exploration and understanding of the debates, especially in small electronic devices, such as mobile phones.

Aiming to exploit the benefits and mitigate the drawbacks of both forums and argumentation tools, in this paper we seek to develop a hybrid approach in which forum-based debates are enriched with argumentative information. In particular, we make use of natural language processing (NLP) and argument mining methods to automatically extract the arguments –together with their constituents and relationships– provided in the citizens' proposals and comments. Hence, instead of requiring users to explicitly express their ideas and opinions as formal argument maps, they freely manifest them as natural language texts.

The proposed approach could have the potential of promoting a higher citizen involvement in public decisions because no technical language and knowledge are necessary to be used in their comments, providing easily understandable arguments in favor and/or against some proposals for public policies. This way, some debates on generally well-known topics could be opened to the opinions of a higher number of citizens, enriching the debate and analysis of the issues they deal with.

Moreover, motivated by the huge use of instant messaging applications, and the little existing research on the potential of such applications for e-participation, the approach proposed in this research is evaluated through a conversational agent or chatbot integrated in the Telegram system¹. In recent years, this type of artificial intelligence agents has been considered for a variety of government-to-citizen scenarios, such as providing automatic attendance on public services [1, 2, 14, 21], easing the access to open government data [11, 27], and supporting citizen consultation [3]. Differently to previous work, this research is framed in a citizen-to-government scenario. We thus envision chatbots as future assistants and even moderators in e-participation processes conducted in online forums, web platforms, and instant messaging applications. In this sense, to the best of our knowledge, incorporating argumentative capabilities to chatbots for exploring user-generated content also represents a novel research contribution.

In brief, this paper presents first steps towards the full implementation of an e-participation chatbot. Leaving as future work the chatbot interaction to moderate participation, we focus on the chatbot capabilities for topic- and argument-driven exploration of citizen-generated content. Specifically, we report an implementation and evaluation of a prototype chatbot for Decide Madrid², the e-participatory budgeting platform of Madrid (Spain), in which residents post, comment and vote ideas and initiatives to improve public services and address existing problems in the city. As an additional differential aspect of our work, the conducted evaluation consists of a user study where both objective and subjective metrics were measured, going beyond the traditional usability and usefulness criteria considered in the literature [11] by assessing the benefits of the proposed argument-driven approach in terms of engagement, persuasiveness, and public values such as transparency and fairness.

The remainder of the paper is organized as follows. Section 2 surveys existing research work on computational argumentation and conversational agents for e-participation. Section 3 presents the proposed argument mining framework, which is composed of an argumentation model and an argument extraction method. Section 4 introduces the developed conversational agent that allows users to access citizen proposals and debates through both topic- and argument-based information exploration mechanisms. Next, Sections 5 and 6 describe the conducted evaluation and discuss the achieved results, respectively. Lastly, Section 7 ends with some conclusions and future research lines.

2 RELATED WORK

We performed a systematic review of the literature on e-participation focused on the use of computational argumentation and conversational agents, since they represent the main topics of interest in our work. For such purpose, we launched formal search queries on the Web of Science³ and Scopus⁴ digital libraries, which index the major journals and conferences in all research fields. Specifically, for computational argumentation in e-participation, we considered a query that retrieves those papers whose title, keywords or abstract contain the regular expression⁵ (“e-part*” OR

¹Telegram instant messaging, <https://telegram.org>

²Decide Madrid e-participatory budgeting platform, <https://decide.madrid.es>

³Web of Science digital library, <https://www.webofscience.com>

⁴Scopus digital library, <https://www.scopus.com>

⁵The asterisk * in the regular expressions means zero or more characters. Hence, for instance, the expression “argument*” is satisfied by the terms *argument*, *arguments*, *argumentation* and *argumentative*, among others.

“*participat*”) AND (“argument*”). For conversational agents in e-participation, we considered an analogous query, but using the regular expression (“e-part*” OR “*participat*”) AND (“conversational” OR “dialog*” OR “chatbot*” OR “chatterbot*” OR “*assistan*”). In the next subsections, we describe the relevant papers obtained from the queries.

2.1 Computational Argumentation in E-participation

With respect to the use of argumentation theories and models in e-participation, the majority of the surveyed work is focused on the development of **computer-supported argumentation visualization** (CSAV) approaches⁶. Instead of unstructured debating, by adding posts of plain text to form conversation threads as done in online forums, collaborative CSAV tools only allow contributions as structured, well-defined elements, such as *issues*, *positions* and *arguments* [25]. Debates are then represented as argument maps, which are visualizations of networks or graphs (generally trees) relating the above elements around a given topic or document. The nodes of a map represent argument constituents (e.g., *claims*, *premises* and *evidences*) and its edges express argumentative relationships (e.g., *support* and *attack*) [7]. A map supports the work of stakeholders by enabling them to manage and navigate through arguments. In this sense, collaborative CSAV tools are aimed to foster a participative process in which the collective effort is put forward in the creation of shared argument maps [17].

Motivated by the information overload problem, in 2009 Loukis et al. [22] proposed the use of CSAV to enhance e-participation. Using the Compendium⁷ tool, they conducted a pilot user survey where 27 people showed moderately positive opinions about the ease of use and usefulness of the proposed argument chart-based visualization, in the context of a debate surrounding a cohabitation law. Contemporaneously with that study, Cartwright and Atkinson [12] investigated the above idea, presenting *Parmenides*, an e-participation prototype system that uses argument schemes to structure and analyze proposals for political action. The authors stated that the system was evaluated positively by students in a user study. However, they did not report details about the conducted experiment.

It is in 2012 when Benn and Macintosh [6, 7] presented *PolicyCommons*, a fully operational platform to facilitate online deliberation of public policies. In addition to issue-centered argument maps, common in CSAV tools, *PolicyCommons* provides a diagram that uses color-coded rectangular blocks to depict the issues within a debate. The size and color of each block reflects the discussion level and topic of the corresponding issue. The authors claimed that the issue map allows addressing the readability and scalability problems. However, they did not report an evaluation of these aspects.

Also in 2012, Panopoulou et al. [25] presented *WAVE*, a web-based CSAV platform developed to facilitate the understanding and debating of the European legislation. Integrated with the *DebateGraph*⁸ tool, *WAVE* allows creating and exploring argument maps, as well as sharing and rating ideas. The authors reported a user study where 319 participants used the platform for discussing about the environment and climate change. Through a questionnaire aimed to evaluate the platform’s ease of use, its facilities to make the discussed topic and underlying debates understandable, and its impact on user participation and engagement, the authors concluded that the platform was in general perceived as valuable and engaging, but generated frustration to some users due to the argument map visualization logic. About 33% of the participants –who mostly were young, educated and computer-literate people– had difficulties to read ideas and navigate through them, especially when the map was too crowded. Also, participants were confused by the lack of a forum-like participation structure.

⁶See [6] for a survey of general-purpose CSAV frameworks.

⁷Compendium - mapping and management of ideas and arguments, <http://compendium.open.ac.uk>

⁸DebateGraph network of thought visualization and sharing tool, <https://debategraph.org>

More recently, in 2018, aimed to empirically measure the impact of the adoption of a collaborative CSAV tool in online political debate with real users, Iandoli et al. [17] reported a user study where a traditional online forum and a collaborative argumentation e-platform were compared. In the study, 95 participants were split into deliberation groups to discuss online reforms of an existing electoral law, using either a conventional forum or a CSAV tool. By analyzing several objective metrics and subjective questionnaire responses related to activity levels, system usability, and quality of collaboration, the authors showed that in the forum, users produced more ideas and activity, and perceived a better quality of the collaboration process, whereas in the CSAV tool, users exchanged more arguments and viewed and rated more posts from others.

Differently to the previous surveyed researches on CSAV, in 2015 Bench-Capon et al. [5] proposed to use computational argumentation –and more specifically, **argument mining**– to elicit justifications of a public policy, and supply critiques for a given proposal and justification. Specifically, they presented an argumentation scheme and a semantic structure for practical reasoning. The scheme considers several policy elements, such as *circumstances*, *goals*, *actions*, *consequences* and *promoted values*. The semantic model is composed of encoded rules that instantiate the scheme on a given domain, and can be interpreted for reasoning arguments, e.g., by logic programming engines. The approach received initial positive user feedback about its capability to support participatory democracy, but requiring a considerable investment of time and expertise for encoding the domain model.

Analyzing the state of the art, we can observe that providing argumentation-supported functionalities in e-participation platforms is generally perceived as valuable by users since they achieve a better understanding of existing proposals and debates, and are capable of getting a better formed opinion and consequently making better decisions. By contrast, as concluded in some of the reported studies, although relevant, the dominant argument map-based visualization makes it difficult to manage and explore the argumentation information.

We thus advocate for maintaining a forum content structure, but enriching it with argument-based information. Attempting to exploit the benefits and mitigate the drawbacks of both forums and CSAV tools, we propose a hybrid approach where users post ideas and comments in free text, and the system automatically extracts and relates the underlying structured arguments. The system then makes use of generated argumentative structures to organize and present the textual content, and allows users to perform an argument-driven information exploration. In this context, as we shall present in subsequent sections, our system manages both topic-centered [6, 7] and issue-centered [17, 25] argument browsing. The system, on the other hand, is built upon a novel, rich argument model with fine-grained argument relationships (e.g., *cause*, *consequence*, *comparison*, *exemplification*) that go beyond the traditional *support-attack* argumentative schema considered in the literature on CSAV for e-participation [17, 25].

2.2 Conversational Agents in E-participation

In the last few years, conversational agents –and more specifically chatbots– are increasingly being considered for a variety of **government-to-citizen (G2C)** applications, such as providing automatic attendance on public services [1, 2, 14, 21], easing the access to open government data [11, 27], and supporting citizen consultation [3].

E-participation –focused on the computer-assisted support to collaborative citizen initiatives– entails other applications for chatbots. Although the idea of incorporating a conversational agent into an e-participation tool was proposed in 2006 by Boden et al. [8], it is in 2019 when the research literature begins to report on the potential of chatbots as facilitators of citizen participation [26, 29], i.e., as **citizen-to-government (C2G)** communication channels.

From a number of interviews with experts, Tavanapour et al. [29] created a list of meta-requirements (e.g., providing answers to topic-related questions, explaining conditions for idea submission) and consequent design principles (e.g.,

providing the capacity to summarize project-related information, handling the phases of the idea generation process) that an e-participation chatbot should have. Taking these requirements and principles into account, they developed a chatbot aimed to assist users with the idea generation task. Among other aspects, the authors highlighted the benefit of using the chatbot to create citizen proposals in a structured and consistent form, an issue that was positively appreciated in a user study (N=32) and may serve as a good starting point for public decision makers to evaluate elaborated proposals. In a related work, Petriv et al. [26] also conducted expert interviews (N=12) with the goal of identifying concerns, limitations and enablers that may affect on the design of chatbots for the public sector. Among others, limited accessibility and lack of technical user skills were stated as major limitations, and positive perception of innovation and provision of public values were considered as principal enablers.

More recently, research work has been published that reports implementations and evaluations of chatbots in e-participation scenarios. Chohan et al. [28] investigated the use of chatbots as a communication interface in citizen science projects that involve citizens and experts in the development of scientific projects and in the use of scientific knowledge to address societal problems. Analyzing the questions given in a user survey (N=13), the authors confirmed that a chatbot can promote participants' motivation, cooperation and engagement within a project. Assuming the need for higher usability and knowledge integration, the authors claimed that chatbots are also beneficial for continuous development, testing and deployment, since they do not rely on complex visual interfaces which may delay research projects and require particular maintenance.

Finally, Haqbeen et al. [15] presented D-Agree, a text-only discussion-processing and decision support platform that aims to harness the wisdom of the crowd for obtaining innovative suggestions that may help policymakers in the development of strategic city plans. The platform is centered around a chatbot that facilitates people to reach agreements during the urban planning processes. More specifically, the chatbot is introduced into online conversations to interact with citizens, moderating discussions by posting facilitated messages and replying to user posts, and encouraging reaching a consensus by mediating and providing arguments given in the posts. In this context, the chatbot asks participants to provide supporting or attacking arguments for posted opinions. An NLP (argument mining) engine is used to automatically classify sentences of a discussion as *issues*, *ideas*, *pros* and *cons*, as well as to extract relationships between sentences. The authors reported a large user study (N=733) analyzing objective activity metrics which evidence the benefits of using the chatbot facilitator to increase participation, promote argued discussion, and achieve higher consensus, allowing the collection of more reliable opinions and the increment of transparency and legitimacy of decision and policy making processes.

As it can be observed in the surveyed literature, conversational agents have been recognized as powerful tools for C2G e-participation applications, but their research is in its infancy. They have been only proposed as discussion facilitators, and have been mainly evaluated in terms of user participation and engagement levels. Based on these pillars, our research proposes a chatbot that aims to support the exploration of citizen-generated content in e-participation platforms, and is evaluated through various metrics, including measures of public value creation. Moreover, similarly to the D-Agree system [15], our chatbot makes use of 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. The argument mining framework of the chatbot is presented next.

3 ARGUMENT MINING FRAMEWORK

In this section, we present our framework to automatically identify arguments in textual content, together with their constituents and relationships. The framework is built upon a generic argumentation model that is based on a rich

Table 1. Types and subtypes of argument relations, with examples of argument linkers for English and Spanish.

Type	Subtype	Intent	English linkers		Spanish linkers	
			no.	examples	no.	examples
Cause	Condition	qualifier	30	if [ever/so], in case of/that	33	si [alguna vez/es así], en caso de/que
	Reason	support	16	because [of], due to, since	16	porque, ya que, debido a [que], pues
			46		49	
Clarification	Conclusion	support	18	to conclude, in/as conclusion	21	para concluir, en/como conclusión
	Exemplification	support	8	for [example/instance], as an example [of]	12	por ejemplo, como ejemplo [de]
	Restatement	support	6	in other words, that is [to say]	26	en otras palabras, es decir, esto es
	Summary	support	12	summarizing, summing up, to sum up	7	resumiendo, concluyendo, para acabar
			44		66	
Consequence	Explanation	support	6	actually, in [actual] fact, indeed	6	realmente, de hecho, en realidad
	Goal	support	19	for, to, in order to, aimed/aiming to	15	para, por, con el fin de
	Result	support	21	therefore, thus, hence, then, so [that]	40	por [lo] tanto, por consiguiente/ende
			46		61	
Contrast	Alternative	support/attack	21	on the other hand, in another case	26	por otra parte, por otro lado, en otro caso
	Comparison	support/attack	7	while, whereas, compared [to/with]	17	mientras [que], comparado con
	Concession	attack	16	although, [even] though, despite [that]	28	aunque, aún/incluso [si/así], a pesar de
	Opposition	attack	22	but, however, nonetheless, albeit	31	pero, sin embargo, no obstante
			66		102	
Elaboration	Addition	support	15	also, besides, as well, too, moreover	17	también, además/aparte [de], [lo que] es más
	Precision	support	11	in particular, particularly, especially	13	en particular, particularmente, especialmente
	Similarity	support	10	similarly/analogously [to], like, likewise	10	similarmente/análogamente [a], como, al igual que
			34		40	
			236		318	

taxonomy of argument types (Subsection 3.1), and consists of an argument extraction method that makes use of natural language processing techniques (Subsection 3.2). For a given text, the framework generates structured information about the extracted arguments (Subsection 3.3).

3.1 Argumentation Model

In the Argument Mining research field [20], most of the existing computational methods and tools to design, extract and share arguments consider *premises* and *claims* as the principal **argumentative units**, and *support* and *attack* (rebuttal) as the possible **argument relations**. Our argument mining framework is built upon this argumentation model, but extends it by including purpose-based types of the above relations. More specifically, we propose a taxonomy with the following argument relation types:

- *Cause*: linking an argument that reflects the *reason* or *condition* for another argument.
- *Clarification*: introducing a *conclusion*, *exemplification*, *restatement* or *summary* of an argument.
- *Consequence*: evidencing an *explanation*, *goal* or *result* of a previous argument.
- *Contrast*: attacking arguments, distinguishing between giving *alternatives*, doing *comparisons*, making *concessions*, and providing *oppositions*.
- *Elaboration*: introducing an argument that provides details about another one, entailing *addition*, *precision* or *similarity* issues about the target argument.

These types of argument relations are automatically identified by an extraction method that makes use of natural language processing techniques. As it will be explained in the next section, the method consists of finding certain *argumentative patterns* within the syntactic trees of sentences, and the patterns are defined around argument *linkers* or connectors, i.e., set of words that link premises and claims of arguments. Table 1 shows examples of English and Spanish linkers for each type and subtype of our argument relation taxonomy. The taxonomy and full lists of linkers are publicly available online⁹.

⁹Argument taxonomy and linkers, <https://github.com/argrecsys>

Algorithm 1 Argument extraction method

```

Require: text, linkerList and syntacticPatternList
argumentList ← createList()
sentences ← splitSentences(text)                                ▷ NLP task
for each sentence s ∈ sentences do                                ▷ NLP task
  tree ← obtainSyntacticTree(s)
  r ← getRoot(tree)
  Q ← createQueue()
  enqueue(Q, r)
  while Q is not empty do                                          ▷ Breadth first search
    u ← dequeue(Q)
    uText ← getNodeText(tree, u)
    if uText ∈ linkerList then
      pattern ← null
      nodes ← getNodesAtLevel(tree, u)
      for each node v ∈ nodes do
        phrasalCategory ← getPhrasalCategory(v)
        pattern ← concat(pattern, phrasalCategory)
      end for
      if pattern ∈ syntacticPatternList then
        claim ← getTextFromNodesOnLeft(tree, u)
        premise ← getTextFromNodesOnRight(tree, u)
        linker ← getLinker(linkerList, uText)
        level ← getNodeDepthLevel(u)
        argument ← createArgument(premise, linker, claim, level, pattern)
        argumentList ← addArgument(argumentList, argument)
      end if
    end if
    children ← getChildren(tree, u)
    for each node v ∈ children do
      enqueue(Q, v)
    end for
  end while
end for
return argumentList

```

3.2 Argument Extraction Method

We propose a heuristic method aimed to automatically identify and extract arguments from textual content, which is evaluated on citizen proposals and comments from the Decide Madrid e-participatory platform. The method searches for certain argumentative patterns in the syntactic trees of input sentences. Such patterns are defined by manual inspection of syntactic phrase structures that have one of the considered linkers (Subsection 3.1) and more frequently appear in a large corpus (from Decide Madrid).

The method follows a simple but effective **algorithm** to address the three basic tasks of argument mining [20], namely *argument detection*, *argument constituent identification* (i.e., *claims* and *premises* related through an argument *linker*), and *argument relation recognition* using the proposed taxonomy (Subsection 3.1).

Table 2. List of the syntactic patterns associated to valid argumentative structures. The patterns are composed of logical combinations of phrasal categories. The phrasal categories are: [conj] = Conjunction, [conj_LNK] = Conjunction that continues with a linker, [grup.verb] = Verb group, [neg] = Negation, [PUNCT] = Punctuation mark, [S] = Clause, [S_LNK] = Clause starting with a linker, [sn] = Noun phrase, [sp] = Prepositional phrase, and [sp_LNK] = Prepositional phrase starting with a linker.

[grup.verb]-[sn]-[S_LNK]*	[sn]-[neg]-[grup.verb]-[S_LNK]*	[sn]-[neg]-[grup.verb]-[sn]-[sp_LNK]*
[neg]-[grup.verb]-[sn]-[S_LNK]*	[sn]-[grup.verb]-[sp_LNK]*	[sp]-[grup.verb]-[sn]-[S_LNK]*
[grup.verb]-[sn]-[sp_LNK]*	[sn]-[neg]-[grup.verb]-[sp_LNK]*	[sp]-[neg]-[grup.verb]-[sn]-[S_LNK]*
[neg]-[grup.verb]-[sn]-[sp_LNK]*	[sn]-[grup.verb]-[sn]-[S_LNK]*	[sp]-[grup.verb]-[sn]-[sp_LNK]*
[S]-[conj_LNK]-[S]*	[sn]-[neg]-[grup.verb]-[sn]-[S_LNK]*	[sp]-[neg]-[grup.verb]-[sn]-[sp_LNK]*
[S]-[conj]-[S_LNK]-[S]*	[sn]-[grup.verb]-[sn]-[sp_LNK]*	[S]-[PUNCT]-[S_LNK]*
[sn]-[grup.verb]-[S_LNK]*		

The algorithm is shown in Algorithm 1. It consists of two consecutive phases, where the outputs of the first phase serve as inputs for the second phase. These phases are *identifying arguments* and *extracting argument constituents and relations*.

3.2.1 Identifying arguments. In this phase, the source text –i.e., a citizen’s proposal or comment– is first split into sentences, where arguments are searched (isolatedly in this stage of our research). Next, the syntactic tree of each sentence is obtained by using the Stanford CoreNLP library¹⁰. Through a breadth first search (BFS), each node of the tree is visited checking the presence or absence of a linker. If the text of a node starts with one of the considered argument linkers, a syntactic structure is constructed by concatenating the phrasal categories of the neighbor nodes of the linker node (including it) according to the language reading order, which is from left to right (in English and Spanish). Once the syntactic structure is constructed, it is compared with each of the valid, manually defined argumentative patterns (Table 2). In case of matching, the argument of the corresponding phrase is extracted –in the next phase– and stored in a temporary structure along with the tree level (depth) where the argument is found. In general, the closer to the tree root an argument is found, the more relevant it is within the sentence.

3.2.2 Extracting argument constituents and relations. Once one of the aforementioned syntactic patterns is matched, the phrase is split into a claim and a premise according to and connected through the linker. The extraction of the claim is performed by concatenating (from left to right) the text of all the sibling nodes before the linker node. The extraction of the premise is performed in the same way, but with the sibling nodes after the linker node. The created argument structure is finally stored into a JSON¹¹ data object whose format is explained next.

3.3 Argumentative Structures

As just mentioned, the extracted arguments are stored in **JSON data objects** for their later use and exploitation. Hence, for the citizen proposal “Allowing pets in public transport”¹², our method automatically identifies and extracts an argument composed of the claim “We are almost forced to use public transport in the city” and the premise “but pets are not allowed in EMT”, which attacks the claim of the argument that supports the proposal (major claim). Figure 1 shows in JSON format a complete argument structure extracted from the above citizen proposal. It contains i) the identifier of the proposal, ii) the sentence where the argument was found, iii) the argument constituents, connector and relation type, subtype and intent, iv) the sentence nouns, verbs, named entities and main verb, and v) its syntactic tree.

¹⁰Stanford CoreNLP library, <https://stanfordnlp.github.io/CoreNLP>

¹¹JavaScript Object Notation (JSON), <https://www.json.org>

¹²Proposal 5717 in Decide Madrid, <https://decide.madrid.es/proposals/5717-permitir-mascotas-en-transporte-publico>

Fig. 1. Example in JSON format of an argument extracted from a citizen proposal about allowing pets in Madrid public transport. EMT stands for “Empresa Municipal de Transportes de Madrid” (i.e., Madrid Regional Transport Company).

```
{
  "5717-1-1": {
    "proposalID": 5717,
    "majorClaim": {
      "entities": "[]",
      "text": "Allowing pets in public transport",
      "nouns": "[pets, transport]"
    },
    "sentence": "We are almost forced to use public transport in the city but pets are not allowed in EMT",
    "claim": {
      "entities": "[]",
      "text": "We are almost forced to use public transport in the city",
      "nouns": "[use, transport, city]"
    },
    "linker": {
      "value": "but",
      "intent": "attack",
      "type": "CONTRAST",
      "subType": "OPPOSITION"
    },
    "premise": {
      "entities": "[EMT]",
      "text": "pets are not allowed in EMT",
      "nouns": "[pets]"
    },
    "mainVerb": "forced",
    "pattern": {
      "value": "[S]-[conj_LNK]-[S]-[PUNCT]",
      "level": 1
    },
    "syntacticTree": "(sentence
      (S (sn (PRP We)) (group.verb (VBP are) ... ))
      (conj but)
      (S (sn (NNS pets)) (group.verb (VBP are) (RB not) ... ))
      (PUNCT .))"
  }
}
```

The extraction of arguments from textual content enables the possibility of finding **argumentative threads** associated to the proposal and its comments. These structures can be interpreted as summaries of conversations aimed at debating certain ideas in favor or against the proposal or some of its aspects. To this end, the proposal description along with the comments –having extracted arguments– can be represented and analyzed as a directed graph where argumentative threads can be found using the longest path algorithm. As an illustrative example, Figure 2 shows an argumentative thread (4 levels deep) extracted from the description and comments of a Decide Madrid proposal¹³ related to the need of a massive tree planting in Madrid.

¹³Proposal 20389 in Decide Madrid, <https://decide.madrid.es/proposals/20389-arborizacion-masiva-en-madrid>

Fig. 2. Argumentative thread extracted from a citizen proposal. C, L and P stand for claim, linker and premise, respectively.

- ```

> Root argument [depth level 0]:
MC: Massive tree planting in Madrid.

- Argument reply [depth level 1]:
C: Planting trees native to the Madrid region.
L: {linker: 'to', intent: 'support', type: 'CONSEQUENCE', subType: 'GOAL'}
P: Improve air quality, maintain a natural lifestyle and improve urban aesthetics with living beings.

- Argument reply [depth level 2]:
C: The first thing they should do is to stop cutting down healthy trees.
L: {linker: 'as', intent: 'support', type: 'CAUSE', subType: 'REASON'}
P: They are doing in Manzanares neighborhood.

- Argument reply [depth level 2]:
C: More than 230 trees in 3 weeks with the excuse that they are very dangerous and will fall on us.
L: {linker: 'but', intent: 'attack', type: 'CONTRAST', subType: 'OPPOSITION'}
P: When they started cutting down, only 4 of the 230 were hollow inside.

- Argument reply [depth level 2]:
C: Then they talk to us about contamination.
L: {linker: 'but', intent: 'attack', type: 'CONTRAST', subType: 'OPPOSITION'}
P: It is a lie, an incongruity and a nonsense.

- Argument reply [depth level 3]:
C: If only the trees they cut down were replaced by younger ones.
L: {linker: 'but', intent: 'attack', type: 'CONTRAST', subType: 'OPPOSITION'}
P: That is not the case.

- Argument reply [depth level 4]:
C: When an old tree falls on people it is a catastrophe.
L: {linker: 'but', intent: 'attack', type: 'CONTRAST', subType: 'OPPOSITION'}
P: We know that for years the care of the trees has not been controlled.

```

## 4 CONVERSATIONAL AGENT

In this section, we present the developed chatbot, which is shown in Figure 3 by means of three screenshots with examples of real user interactions. Next, we describe the chatbot supported conversation intents (Subsection 4.1) and its high-level architecture (Subsection 4.2).

### 4.1 Conversation Intents

In modern conversational agents and chatbots, such as those developed through technologies like Google Dialogflow<sup>14</sup>, IBM Watson<sup>15</sup>, and Microsoft LUIS<sup>16</sup>, a conversation is usually composed of **intents** that represent different user information needs (purposes or goals). An intent can be independent of the rest of intents, or should only be considered after addressing another particular intent.

<sup>14</sup>Google Dialogflow, <https://cloud.google.com/dialogflow>

<sup>15</sup>IBM Watson, <https://www.ibm.com/watson>

<sup>16</sup>Microsoft LUIS, <https://www.luis.ai>

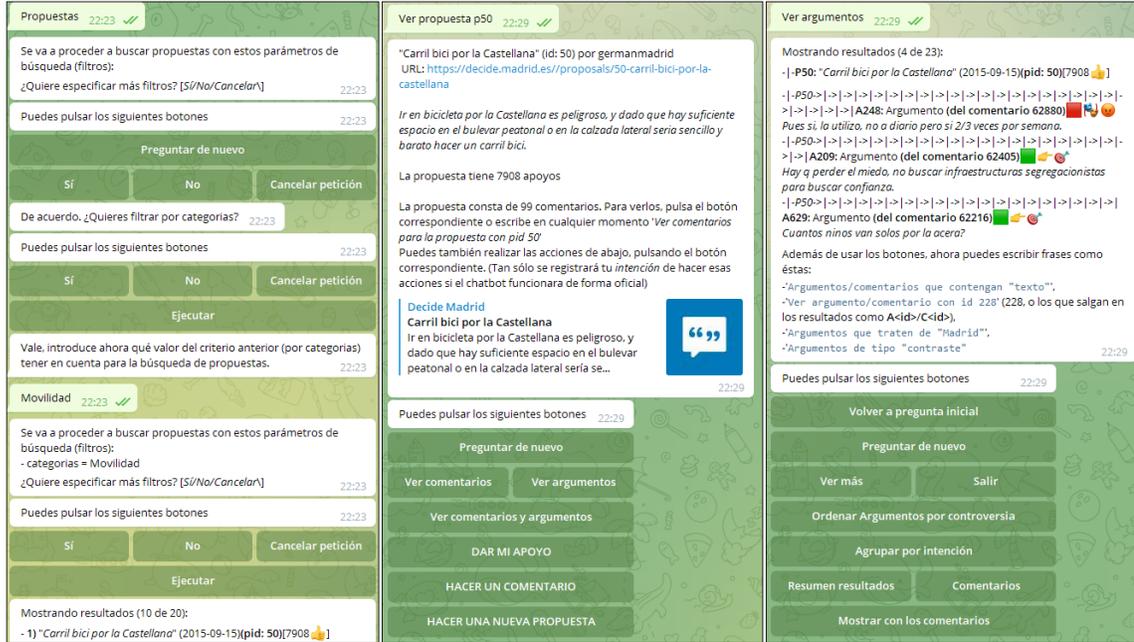


Fig. 3. Three screenshots of human conversations with the chatbot. From left to right, they show a filtering process of proposals, details of a given proposal, and a set of categorized arguments existing in a proposal’s comments. In the latter case, the intent, type and subtype of each argument are depicted through representative emojis.

During a conversation, an intent is triggered when the user produces an utterance that satisfies a particular sentence pattern, which has been defined in advance in the chatbot NLP modeling process. Appropriate responses (i.e., natural language messages and menus for interaction) are generated for each utterance and intent.

A possible feature of a chatbot is the storing of log records with the user’s actions, queries and feedback, along with associated timestamps and annotations. This functionality was included in our chatbot to measure various performance metrics during the experiments (Section 5).

Figure 4 shows a diagram with the conversation intents handled by our chatbot. In the following, we describe the most relevant ones, and provide some of their input (triggering) sentence patterns and output results and implications.

- *Welcome*. This intent is triggered automatically at the beginning of a conversation and also by greeting the chatbot. In it, the chatbot welcomes and offers its help to the user.
- *Help*. In this intent, the chatbot provides an exhaustive description of its functionalities. It is triggered when the user introduces sentences like “I need some advice,” “Can you help me?” or simply “Help.” The help documentation is distributed through several themes, accessible by buttons, and associated to the entities retrievable by the chatbot, namely proposals, comments and arguments.
- *Categories*. This intent is aimed to iteratively show the list of available proposal categories. It is triggered with sentences like “Categories” and “What categories are available?”
- *Topics*. This intent is aimed to iteratively show the list of available proposal topics, each of them belonging to certain category. It is triggered with sentences like “Topics” and “Topics of the category urbanism.”

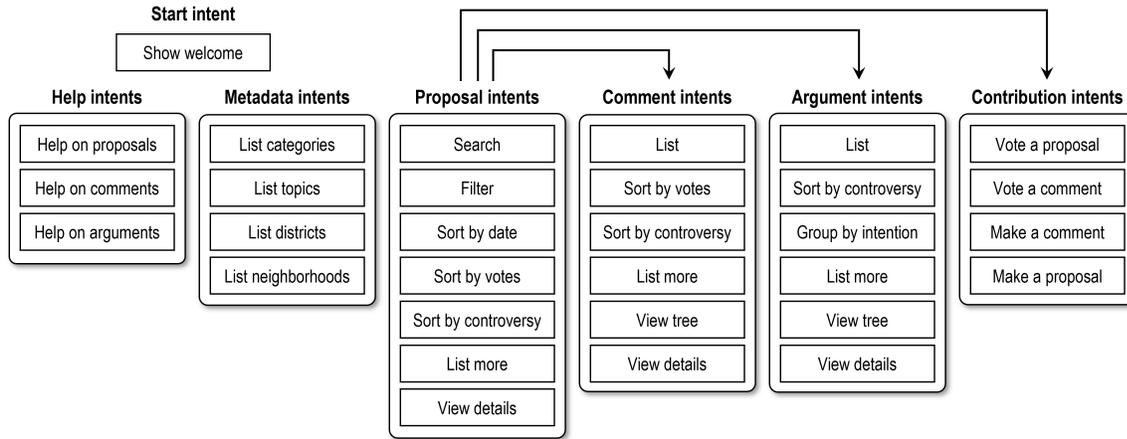


Fig. 4. Conversation intents of the chatbot.

- Proposals.** This intent allows the user to search for citizen proposals. It is triggered when the user writes sentences like “Proposals” or “Proposals of the category mobility”, or even “Proposals of the category mobility, topic taxis in the district of Latina and neighborhood of Aluche.” The latter utterance example contains all available options for filtering requested proposals: *by category*, *by topic*, *by district* and *by neighborhood*. These filters do not have to be specified all at once, as the chatbot will ask the user for them iteratively (see left screenshot in Figure 3). In any case, the retrieved proposals can be sorted in three different ways: *by date*, *by number of votes* and *by controversy* (as computed in [10]).
- Details of a proposal.** 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, which appeared shortened in proposal lists. Examples of sentences that trigger this intent are: “Proposal with id 7” and “Details of proposal 891.” The data presented for a proposal include its title, summary (used as the *major claim* for possible arguments), and number of votes (see middle screenshot in Figure 3). The chatbot also provides some buttons that allow the user to access more comments and arguments of the proposal, and give several types of feedback: *vote the proposal*, *make a comment*, and *create a new proposal*.
- Comments.** This intent can be triggered at any time when sentences like “Show comments of proposal with id 7234” are introduced. Also, if a proposal search was recently executed, utterances like “Comments from last proposals” are recognized. The intent allows exploring iteratively all the comments of a given proposal or list of proposals. Since comments are not only issued against proposals, but against other comments, the chatbot also allows exploring the underlying *tree* of comments (debate). A simplification of this structure is provided to the user if she chooses to see summary of results, like “most commented proposal/comment” and “number of debate threads” in the tree. The intent can also be accessed through buttons in the *view details* response.
- Arguments.** This intent offers analogous functionalities to the *Comments* intent, but applied to arguments extracted from proposal comments (see right screenshot in Figure 3). In this case, an additional type of utterance is allowed, which is *group arguments by intent* (topic and subtopic). As for comments, either the tree or statistics of arguments of a proposal can be visualized, by pressing the corresponding buttons provided by the chatbot.

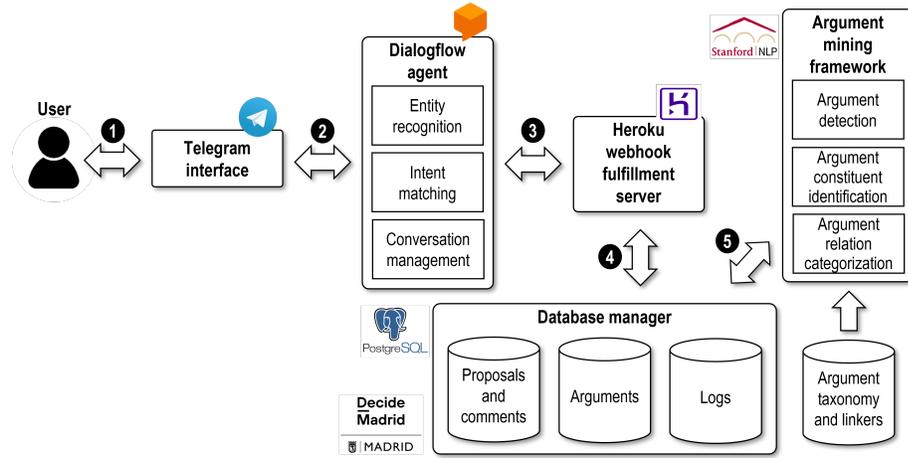


Fig. 5. Architecture of the chatbot.

## 4.2 Architecture

Figure 5 shows the **high-level architecture** of the developed chatbot, which is built upon the Google Dialogflow framework. This framework enables relatively easy and fast implementations of conversational agents by processing user utterances in natural language, managing modeled human-machine conversations, handling the connection to external services and data sources, and integrating the communication with commercial instant messaging and social networking applications, such as Google Assistant, Facebook Messenger, WhatsApp, Telegram and Skype. Our chatbot was integrated with the Heroku cloud computing platform for accessing internal PostgreSQL databases, and with the Telegram application for communicating with the user.

The figure depicts the main modules of the chatbot and enumerates the steps of the data flows between the modules. As just mentioned, Telegram is used as the user interface to communicate with the chatbot (step 1 in the figure). However, thanks to the Dialogflow framework, the chatbot could be easily adapted to be used in other applications. In addition to manage the transmission of input and output messages with Telegram (step 2), a Dialogflow agent is capable of several natural language processing tasks, such as automatic recognition of entities (e.g., names of people and places, dates, numbers) in input sentences, matching user utterances with previously defined conversation intents, and managing the flow of conversations, i.e., triggering of events associated to intents and transitions between intents. The definition of intents is done through the online platform of Dialogflow and involves, among other things, the provision of examples of sentences and their corresponding entities. By means of deep learning methods, the agent is capable of recognizing patterns in (new) equivalent input sentences.

When a conversation intent is triggered during a conversation, the Dialogflow agent calls a remote web service which is in charge of executing certain logic of the system. This is done through the Heroku webhook fulfillment server (step 3). The services access the chatbot databases, which are in a remote server and gather the citizen proposals and comments from Decide Madrid, the automatically extracted arguments, and the logs recording the users' activity on the chatbot. The data access is controlled by an ad hoc manager (step 4). The databases are also accessed offline by the argument mining framework (step 5), which, as explained in Section 3, makes use of the CoreNLP library for detecting arguments in text, identifying argument constituents (i.e., claim, link, premise), and categorizing argument relations.

## 5 EXPERIMENTS

In this section, we present the empirical experiment performed to evaluate the developed chatbot. We briefly describe the case study considered for the experiment (Section 5.1), the addressed research goals and hypotheses (Section 5.2), the used evaluation methodology (Section 5.3) and metrics (Section 5.4), and the participants of the experiment (Section 5.5).

### 5.1 Case Study

We have implemented and evaluated our chatbot with publicly available citizen-generated content from the **Decide Madrid**<sup>17</sup> **e-platform**. This platform is an ad-hoc website used by the City Council of Madrid (Spain) as part of its participatory budgeting initiative since September 2015. Through the tool, residents of Madrid can post proposals to address issues and problems in the city, and comment and vote others' proposals. On a yearly basis, those citizen proposals that receive certain number of votes are technically and economically assessed, and eventually are funded and implemented by the city government. The budget allocated to these proposals was 50 million euro in 2021.

Similarly to [10], the selection of Decide Madrid as a representative e-participation tool was motivated by two reasons. First, participatory budgeting (PB) is among the most widely used citizen participation methods worldwide. From a total of over 1,900 citizen participation cases reported in Participedia.net, around 600 cases correspond to PB initiatives<sup>18</sup>. Also, according to the Participatory Budgeting Project<sup>19</sup>, more than 7,000 cities around the world have implemented PB processes. Second, Decide Madrid follows a standard structure and architecture of electronic PB tools (e.g., Stanford Participatory Budgeting<sup>20</sup> and EU Open Budgets<sup>21</sup>), where web pages held proposals descriptions and metadata, as well as debates and supports on proposals. In fact, Decide Madrid is implemented upon the CONSUL<sup>22</sup> open-source framework, which as far of January 2022 has been utilized by 135 institutions of 35 countries supporting 90 million citizens.

More specifically, we instantiated our chatbot with the Decide Madrid open dataset previously used and publicly provided in [10]. The dataset contains information about 21,744 citizen proposals –automatically classified into 30 categories and 325 topics, geolocated in 21 city districts (and many of them in 129 city neighborhoods), and annotated with controversy scores–, and 62,838 comments –automatically processed with the argument mining framework presented in Section 3. To narrow the scope of the user study, we limited the use of the chatbot to a subset of 80 proposals and their associated 5,633 comments. We provide the dataset and the source code of the argument mining framework and chatbot in <https://github.com/argrecsys>.

### 5.2 Research Goal and Hypotheses

In addition to **evaluating to what extent a conversational agent or chatbot can be an appropriate tool in an e-participation context**, we also aimed to assess the **benefits of using argument-driven information exploration in e-participation** with respect to a traditional topic keyword-based navigation. For such purpose, in the study, 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.

<sup>17</sup>Decide Madrid e-participatory budgeting platform, <https://decide.madrid.es>

<sup>18</sup>Participatory budgeting cases in Participedia.net, <https://participedia.net/search?selectedCategory=case&query=%20budgeting>

<sup>19</sup>Participatory Budgeting Project, <https://www.participatorybudgeting.org>

<sup>20</sup>Stanford Participatory Budgeting, <https://pbstanford.org>

<sup>21</sup>EU Open Budgets, <https://openbudgets.eu>

<sup>22</sup>CONSUL open-source citizen participation framework, <https://consulproject.org>

More specifically, inspired by previous work (e.g., [11]), in our study we stated the following research hypotheses associated to potential citizen participation- and public value-related benefits:

- **H1 (usability)**: The users of the argument-driven chatbot achieve a better understanding of the citizen proposals and their pros and cons.
- **H2 (usefulness)**: The users of the argument-driven chatbot perceive the system as more valuable for getting well-formed opinions and making better decisions in the participatory process.
- **H3 (persuasiveness)**: The users of the argument-driven chatbot are more willing to rethink their initial points of view, or to make own proposals and comments.
- **H4 (transparency)**: The users of the argument-driven chatbot feel that using the system are more able to explore a representative sample of citizen proposals and debates.
- **H5 (fairness)**: The users of the argument-driven chatbot feel that the system is inclusive and provides access to heterogeneous ideas and comments, even those that are related to controversial issues, or affect to minority or discriminated groups.
- **H6 (satisfaction)**: The users of the argument-driven chatbot are more satisfied with the information search and exploration functionalities of the system.
- **H7 (engagement)**: The users of the argument-driven chatbot perform more activities, such as access and support to citizen ideas and comments.

To validate these hypotheses, we analyzed a number of objective and subjective metrics measured during the study, as explained in Sections 5.3 and 5.4. Similarly to past research on e-participation chatbots (Section 2.2), in addition to the above benefits and metrics, we also took several metrics into account for measuring the perceived **ease of use**, **effectiveness** and **efficiency** of the chatbot.

### 5.3 Evaluation Methodology

The design of our user study was done considering evaluations of chatbots for the public sector reported in the research literature [2, 11, 26]. In our case, to make the study as realistic as possible, it was conducted in an **uncontrolled setting** where, without external supervision, participants freely tested the chatbot via Telegram during a period of one week. They used their own Telegram accounts and mobile devices, having similar computing capabilities and internet connection conditions. All their interactions with the chatbot were monitored and recorded anonymized in a database.

Differently to previous empirical experiments where participants focused on a single law [17, 22] or political issue [25], in our study, the users were allowed to explore all the citizen proposals and comments of the chatbot database, with no particular task requested. The users in the *experimental group* also had access to the arguments automatically extracted from the proposals descriptions and debates.

Participants were recruited forming a heterogeneous set of people with different demographic attributes (i.e., age and gender), educational backgrounds, and knowledge levels about chatbots and citizen participation. Before using the chatbot, participants filled a consent form where they provided the above personal data. They also received a few instructions concerning the use of the chatbot and the help documentation available in the chatbot. Finally, after testing the chatbot, participants filled a questionnaire aimed to capture their opinions about the potential user participation- and public value-related benefits of the chatbot presented in Section 5.2.

Both objective interaction records and subjective opinions were considered as metrics to evaluate the chatbot, as explained in the next section.

Table 3. Items of the developed questionnaire to evaluate the proposed online metrics.

| Criterion      | Questionnaire item                                                                                                |
|----------------|-------------------------------------------------------------------------------------------------------------------|
| Ease of use    | I1: The chatbot is easy to use                                                                                    |
|                | I2: The interaction with the chatbot does not require a lot of mental effort                                      |
|                | I3: The help documentation of the chatbot is easy to understand                                                   |
|                | I4: The help documentation of the chatbot is complete                                                             |
|                | I5: The help documentation of the chatbot is valuable                                                             |
| Effectiveness  | I6: The chatbot understands the user’s questions and commands                                                     |
|                | I7: The chatbot gives correct responses to the user’s requests                                                    |
| Efficiency     | I8: The chatbot is ready to interact soon after invocation                                                        |
|                | I9: The chatbot provides responses quickly                                                                        |
| Usability      | I10: The chatbot allows exploring the citizen proposals about certain topic                                       |
|                | I11: The chatbot allows exploring the content of a citizen proposal                                               |
|                | I12: The chatbot allows exploring the pros and cons of a citizen proposal                                         |
| Usefulness     | I13: The chatbot allows finding out the city problems and citizens’ concerns                                      |
|                | I14: The chatbot allows understanding others’ ideas and opinions about citizen proposals                          |
|                | I15: The chatbot allows getting well-formed opinions and making better decisions in the participatory process     |
| Persuasiveness | I16: The chatbot promotes rethinking initial opinions about citizen proposals                                     |
|                | I17: The chatbot promotes commenting on citizen proposals                                                         |
|                | I18: The chatbot promotes making own proposals for the city                                                       |
| Transparency   | I19: The chatbot allows exploring a representative sample of citizen proposals                                    |
|                | I20: The chatbot allows exploring a representative sample of citizen comments and opinions in the citizen debates |
| Fairness       | I21: The chatbot allows exploring an unbiased sample of citizen proposals                                         |
|                | I22: The chatbot allows exploring an unbiased sample of citizen comments (opinions) in the debates                |
|                | I23: The chatbot allows getting informed about controversial issues in the city                                   |
|                | I24: The chatbot allows getting informed about city issues affecting to minority or discriminated groups          |
| Satisfaction   | I25: I am satisfied with the functionalities provided by the chatbot                                              |
|                | I26: I am satisfied with the interaction (communication) offered by the chatbot                                   |
|                | I27: I am satisfied with the current version of the chatbot                                                       |
| Engagement     | I28: I liked using the chatbot as a citizen participation tool                                                    |
|                | I29: I enjoyed using the chatbot                                                                                  |
|                | I30: I would use the chatbot again                                                                                |
|                | I31: I would recommend the chatbot to other people                                                                |
|                | I32: I am going to enter into the Decide Madrid platform                                                          |
|                | I33: I am going to search for information about (electronic) citizen participation initiatives                    |

#### 5.4 Evaluation Metrics

Similarly to [11, 17], 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 logs were used to measure a variety of metrics related to the users’ activity and engagement on the chatbot. Some of these metrics were *usage time*, *number of browsed (lists of) proposals/comments/arguments*, *number of provided votes/comments*, and *number of manifested intentions to create new proposals*.

Regarding the **online evaluation**, at the end of the testing phase, participants filled a questionnaire aimed to measure the system performance, citizen participation, and public value criteria presented in Section 5.2. The questionnaire was composed of thirty three items on a 5-point Likert scale –from strongly disagree (1) to strongly agree (5)–, which are shown in Table 3. In addition to items used in [17, 25], which were constructed on the basis of several theories and studies, as a novel contribution, some of our items attempted to capture the perceived understanding of the citizen proposals and underlying debates, and the perceived utility of the system to support personal decision making.

Table 4. Statistics about participants' activity on the chatbot without and with arguments.

|                                                       | control group | experimental group |
|-------------------------------------------------------|---------------|--------------------|
| <i>avg. number of sessions per user</i>               | 2.8           | 2.8                |
| <i>avg. duration of session (in minutes)</i>          | 16.0          | 23.3               |
| <i>avg. number of actions per user</i>                | 56.8          | 64.9               |
| <i>ask for help</i>                                   | 13.5          | 10.8               |
| <i>list categories/topics/districts/neighborhoods</i> | 6.4           | 7.2                |
| <i>filter proposals</i>                               | 8.0           | 15.6               |
| <i>sort proposals</i>                                 | 2.7           | 2.1                |
| <i>explore proposals</i>                              | 11.3          | 9.8                |
| <i>explore comments</i>                               | 7.6           | 6.7                |
| <i>explore arguments</i>                              | -             | 7.4                |
| <i>provide feedback (new vote/comment/proposal)</i>   | 1.7           | 2.1                |

In addition to these items, the questionnaire also had three open questions where participants were asked to express benefits and/or best features of the chatbot, its drawbacks and/or worst features, and the pros and cons of the chatbot with respect to a traditional web forum.

## 5.5 Participants

A total of 32 people participated in our study. In particular, they were 22 male and 10 female of ages ranging 18-29 years old (12), 30-39 years old (9), 40-49 years old (5), 50-59 years old (4), and more than 59 years old (2), with different education levels: secondary education (3), vocational education (1), Bachelor's degree (20), Master's degree (6), and Doctoral degree (2). Those with Higher Education levels had studied Sciences (2), Social Sciences (8), Arts and Humanities (4), and Engineering (11) careers. Finally, participants had relatively low levels of knowledge/expertise on chatbots –null knowledge and expertise (5), null expertise (5), low expertise (20), and medium expertise (2)– and on citizen participation –null (7), low (16) and medium (9). Previously to the study, 21 participants did not know Decide Madrid, 6 participants were aware but had not used the platform, and only 5 had visited it.

## 6 RESULTS

In this section, we report and analyze the results of the offline and online evaluations introduced in Section 5. We remind that participants were randomly and uniformly split into two groups: a *control group* that utilized the chatbot without the argument-driven functionalities enabled, and an *experimental group* that utilized a full instantiation of the chatbot. In both groups, participants were allowed to freely utilize the chatbot within a period of one week. 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.

Table 4 shows some statistics about the **activity on the chatbot** performed by participants from each group and recorded in the logs database. As it can be seen, our hypotheses were satisfied. Although there was no significant difference on the average number of sessions<sup>23</sup> 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 increment of 45.6% on the average session duration (from 16.0 to 23.3 minutes).

<sup>23</sup>The sessions of a particular user were established by sorting and grouping her log records so that consecutive timestamps differ in at most 15 minutes. Thus, two consecutive logs with timestamps differing in more than 15 minutes were considered as belonging to two different sessions. Only a few cases close to 15 minutes occurred, and no significant result differences were shown by considering other time threshold.

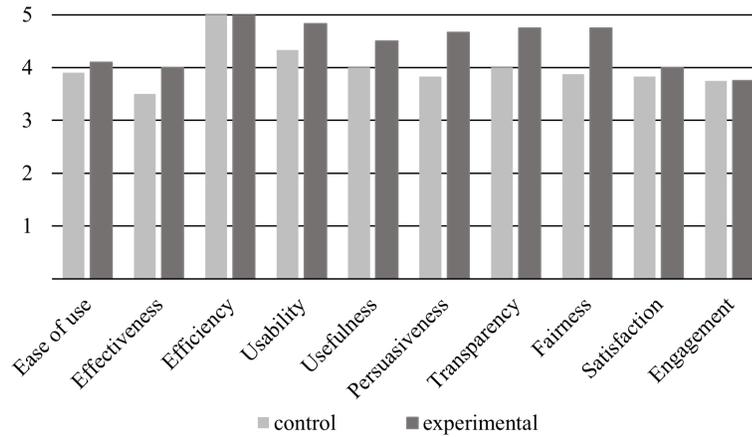


Fig. 6. Average of score medians given by participants in the questionnaire items grouped by evaluation criterion, for the chatbot without/with argument-driven information exploration.

This does not seem to be due to a higher difficulty of use of the full chatbot if we take the number of actions into account. As it can be seen in the table, there was an increment of 14.3% (from 56.8 to 64.9) on the average number of actions per user, and an increment of 23.5% (from 1.7 to 2.1) on the average number of feedback provision actions per user, i.e., expressing intention for creating a new vote, comment or proposal. This result also evidences **higher user persuasiveness and engagement** in the experimental group. Moreover, the average number of actions related to the exploration of arguments (7.4) is meaningful; people in that group preferred to inspect the argument trees rather than the comment threads. We thus claim first insights about the fact that checking the arguments given by citizens entails an increase of participation and involvement.

Figure 6 shows a summary of the 5-scale scores given by participants in the **opinion questionnaire** about the ten chatbot evaluation criteria presented in Section 5.2; specifically, it shows the averages of score medians in each evaluation criterion. Three of these criteria are related with the system performance in terms of ease of use, effectiveness and efficiency, whereas the remaining seven are related with citizen participation and public value benefits: usability, usefulness, persuasiveness, transparency, fairness, satisfaction and engagement. Each criterion was measured through a set of items in the questionnaire (see Table 3). Due to lack of space, the figure only shows the average values of the items' score medians in each set, rather than the median values of each item.

As it can be observed, there were not significant differences between the control and experimental groups with respect to the perception of ease of use and efficiency of the chatbot. Participants found the chatbot **moderately easy to use** (giving average median scores of 3.9 and 4.1 in control and experimental groups, respectively) and **highly efficient** (5.0 in both groups). Regarding effectiveness, the members of the experimental group found the responses given by the chatbot as more accurate (3.5 vs. 4.0). This may be due to a perceived value of the provided argumentative information.

More important differences were obtained in the three levels of 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 the three evaluation criteria, the participants of 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 is observed on the perceived levels of **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. According to these and the previous results, we can claim that having argument browsing functionalities plays a relevant role to promote citizen participation and public values.

Finally, **satisfaction** and **engagement** were equally and positively evaluated in the two versions of the chatbot. However, they obtained more moderate score values in comparison to other evaluation criteria. In the open responses to the questionnaire, participants expressed some limitations and weaknesses of the chatbot for which higher satisfaction scores would not be given. We next briefly present the most frequent positive and negative opinions reflected in the questionnaire.

Among the chatbot features positively evaluated by participants, two stand out: its efficiency and its summarization capability. Users appreciated the fast way to ask for and obtain information through the chatbot, and the direct and compact presentation of such information. Participants also highlighted the transparency and lack of bias on the information presented by the chatbot, since they could check arguments in favor and against proposals in a structured way, and organize comments and arguments by controversy. In general, users stated that the chatbot was easy to use once its commands were learned. To achieve this, they appreciated the detailed and clear help documentation provided by the chatbot.

The errors and complications occurred when input user utterances were not understood by the chatbot represent the principal and more generalized complaint of participants, who expressed the need for more flexibility on the chatbot commands; in particular, they requested support for a less strict syntax. On the opposite communication direction, some participants also missed a more “natural” conversation, i.e., a more colloquial language by the chatbot. Finally, as more specific (technical) issues, there were suggestions to make a more fluent transition between browsed proposals and to facilitate the reading of proposals with large descriptions.

Comparing the chatbot with a traditional web forum, some participants found the chatbot as a better system to search for information, since it offers a direct and fast access to concise content. Two participants also found the chatbot as more accessible due to its adaptation to mobile devices. By contrast, some users stated that the chatbot entails more effort (actions) to navigate through the conversation threads and difficulties to contribute to existing debates.

## 7 CONCLUSIONS

In this paper, we have empirically investigated two promising research lines in e-participation: the use of conversational agents or chatbots as citizen-to-government communication channels, and the exploitation of argument mining techniques to automatically extract and present argumentative information from citizen-generated content.

Specifically, through a user study (N=32) we have evaluated a prototype chatbot that enables a rich, interactive exploration of citizen proposals and debates existing in a real e-participatory budgeting platform. Among other functionalities, the chatbot allows a user to access structured and linked arguments given in favor and against the platform proposals.

The results achieved in our experiments represent first insights about the benefits of the proposed solution in terms of various citizen participation and public value criteria. In addition to facilitating a better search and exploration of the content, pros and cons of proposals about certain topic, the chatbot also helps on finding out and understanding city problems and citizens’ concerns, and consequently on getting well-formed opinions for making better decisions in participatory processes. In addition, the provision of argumentative information entails a greater user perception of transparency and fairness.

The results, on the other hand, have evidenced limitations and weaknesses of the chatbot that have to be addressed in the future. First, there is the need for providing more flexibility on the natural language formulation of user utterances, as well as more colloquial language in the conversation responses from the chatbot. Second, there are suggestions for improving the visualization of large texts and the navigation from one to another proposal.

Apart from these issues, we envision the opportunity of extending the chatbot in several directions. We could incorporate personalized recommendation mechanisms to proactively present information to the user, thus mitigating the information overload problem. We could also develop richer data structures, analysis and visualizations for facilitating decision making. Moreover, we could implement functionalities oriented to citizen collaboration; to the best of our knowledge, the addition of comments into the debates and the creation of new proposals are challenging tasks that have not been addressed by chatbots yet. Finally, we could investigate the integration of the chatbot with external data sources, such as open government data collections and news items, which may be used to complement citizen proposals and verify associated arguments.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation (PID2019-108965GB-I00), the Regional Government of Andalusia (P20\_00314 and B-SEJ-556-UGR20), and the Centre of Andalusian Studies (PR137/19). The authors thank to all people who participated in the reported study.

## REFERENCES

- [1] Aggeliki Androutsopoulou, Nikos Karacapilidis, Euripidis Loukis, and Yannis Charalabidis. 2019. Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly* 36, 2 (2019), 358–367.
- [2] Naomi Aoki. 2020. An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly* 37, 4 (2020), 101490.
- [3] Shubham Atreja, Pooja Aggarwal, Prateeti Mohapatra, Amol Dumrewal, Anwesh Basu, and Gargi B Dasgupta. 2018. Citicafe: An interactive interface for citizen engagement. In *23rd International Conference on Intelligent User Interfaces*. ACM, 617–628.
- [4] Cheuk Hang Au and Kevin KW Ho. 2019. Deliberation in mobile messaging application: A case in Hong Kong. *Communications of the Association for Information Systems* 45, 1 (2019), 17.
- [5] Trevor Bench-Capon, Katie Atkinson, and Adam Wyner. 2015. Using argumentation to structure e-participation in policy making. In *Transactions on Large-Scale Data- and Knowledge-Centered Systems XVIII*. Springer, 1–29.
- [6] Neil Benn and Ann Macintosh. 2011. Argument visualization for eParticipation: Towards a research agenda and prototype tool. In *Proceedings of the 3rd IFIP WG 8.5 International Conference on Electronic Participation*. Springer, 60–73.
- [7] Neil Benn and Ann Macintosh. 2012. PolicyCommons – Visualizing arguments in policy consultation. In *Proceedings of the 4th IFIP WG 8.5 International Conference on Electronic Participation*. Springer, 61–72.
- [8] Cordula Boden, Jessika Fischer, Kathrin Herbig, and Ulrike Spierling. 2006. CitizenTalk: Application of chatbot infotainment to e-democracy. In *Proceedings of the 3rd International Conference on Technologies for Interactive Digital Storytelling and Entertainment*. Springer, 370–381.
- [9] Enrique Bonsón, Sonia Royo, and Melinda Ratkai. 2015. Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly* 32, 1 (2015), 52–62.
- [10] Iván Cantador, María E Cortés-Cediel, and Miriam Fernández. 2020. Exploiting Open Data to analyze discussion and controversy in online citizen participation. *Information Processing & Management* 57, 5 (2020), 102301.
- [11] Iván Cantador, Jesús Viejo-Tardío, María E Cortés-Cediel, and Manuel Pedro Rodríguez Bolívar. 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*. ACM, 168–179.
- [12] Dan Cartwright and Katie Atkinson. 2009. Using computational argumentation to support e-participation. *IEEE Intelligent Systems* 24, 5 (2009), 42–52.
- [13] Viktor Chagas, Isabele Mitozo, Samuel Barros, João Guilherme Santos, and Dilvan Azevedo. 2021. The 'new age' of political participation? WhatsApp and call to action on the Brazilian senate's consultations on the e-cidadania portal. *Journal of Information Technology & Politics* (2021), 1–16.
- [14] Sohail Raza Chohan, Guangwei Hu, Asad Ullah Khan, Ahmad Tisman Pasha, and Muhammad Atif Sheikh. 2021. Design and behavior science in government-to-citizens cognitive-communication: a study towards an inclusive framework. *Transforming Government: People, Process and Policy* (2021).

- [15] Jawad Haqbeen, Sofia Sahab, Takayuki Ito, and Paola Rizzi. 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, 10 (2021), 5453.
- [16] Teresa M Harrison, Theresa A Pardo, and Meghan Cook. 2012. Creating open government ecosystems: A research and development agenda. *Future Internet* 4, 4 (2012), 900–928.
- [17] Luca Iandoli, Ivana Quinto, Paolo Spada, Mark Klein, and Raffaele Calabretta. 2018. Supporting argumentation in online political debate: Evidence from an experiment of collective deliberation. *New Media & Society* 20, 4 (2018), 1320–1341.
- [18] Paul A Kirschner, Simon J Buckingham-Shum, and Chad S Carr. 2012. *Visualizing argumentation: Software tools for collaborative and educational sense-making*. Springer Science & Business Media.
- [19] Daniel Lathrop and Laurel Ruma. 2010. *Open government: Collaboration, transparency, and participation in practice*. O'Reilly Media, Inc.
- [20] John Lawrence and Chris Reed. 2020. Argument mining: A survey. *Computational Linguistics* 45, 4 (2020), 765–818.
- [21] Andreas Lommatzsch. 2018. A next generation chatbot-framework for the public administration. In *18th International Conference on Innovations for Community Services*. Springer, 127–141.
- [22] Euripidis Loukis, Alexandros Xenakis, and Nektaria Tseperli. 2009. Using argument visualization to enhance e-participation in the legislation formation process. In *Proceedings of the 1st International Conference on Electronic Participation*. Springer, 125–138.
- [23] Ann Macintosh. 2004. Characterizing e-participation in policy-making. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. IEEE, 10–pp.
- [24] Ann Macintosh, Thomas F Gordon, and Alastair Renton. 2009. Providing argument support for e-participation. *Journal of Information Technology & Politics* 6, 1 (2009), 43–59.
- [25] Eleni Panopoulou, Efpraxia Dalakiouridou, Efthimios Tambouris, and Konstantinos Tarabanis. 2012. Citizens' evaluation of an online argument visualisation platform for eParticipation. In *Proceedings of the 4th IFIP WG 8.5 International Conference on Electronic Participation*. Springer, 49–60.
- [26] Yulia Petriv, Regina Erlenheim, Valentyna Tsap, Ingrid Pappel, and Dirk Draheim. 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*. Springer, 320–335.
- [27] Simone Porreca, Francesco Leotta, Massimo Mecella, Stavros Vassos, and Tiziana Catarci. 2017. Accessing government open data through chatbots. In *17th International Conference on Web Engineering*. Springer, 156–165.
- [28] Manuel Portela. 2021. Interfacing participation in citizen science projects with conversational agents. *Human Computation* 8, 2 (2021), 33–53.
- [29] Navid Tavanapour, Mathis Poser, and Eva A C Bittner. 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*.
- [30] María Rosalía Vicente and Amparo Novo. 2014. An empirical analysis of e-participation. The role of social networks and e-government over citizens' online engagement. *Government Information Quarterly* 31, 3 (2014), 379–387.