



Designing a conversational agent for supporting data exploration in citizen science

Carolin Stein^{1,2} · Timm Teubner^{3,4} · Stefan Morana⁵

Received: 31 July 2023 / Accepted: 14 February 2024
© The Author(s) 2024

Abstract

Data is ubiquitous in today's digitized society. However, access to and literacy in handling data plays a pivotal role in determining who can benefit from it and who can use—or potentially misuse—it. To combat inequalities and address issues such as misinformation, it is essential to enable citizens to effectively access and understand data within their local ecosystems. To address this challenge, we focus on the case of citizen science and propose using a conversational agent to support data exploration and lower barriers to citizen engagement in research projects. Using a design science research approach, we derive design principles and develop a prototypical artifact. Moreover, we conduct an experimental evaluation, demonstrating strong interest among citizens to participate in scientific data analysis and that conversational agents hold great potential in increasing data literacy.

Keywords Conversational agents · Citizen science · Design science research

JEL Classification O35 · O36 · I24

Introduction

The emergence of the digital era has undeniably amplified the profound impact of data on all aspects of our lives. Technological advancements have enabled the collection and

storage of large datasets (Clarke, 2016; Twidale et al., 2013). Data-driven business models motivate companies to collect and analyze increasing amounts of data, in some cases even at the expense of their customers' interests (Trzaskowski, 2022). While the increased availability of data can lead to new insights and better decisions, it also accentuates issues of inequality and exploitation (D'Ignazio, 2022). Ownership and literacy of data are critical factors in determining who can effectively utilize data to their advantage (D'Ignazio, 2022). This becomes particularly concerning as data and its products can be misused for personal, political, or economic reasons (Carmi et al., 2020; Pullinger, 2021; Trzaskowski, 2022). The evolving landscape of generative AI poses further challenges with the proliferation of disinformation in the digital realm (Hanley & Durumeric, 2023). In this context, data literacy becomes essential, not just for actively engaging in public debates and decision-making (Debruyne et al., 2021; Radermacher, 2021; Schüller et al., 2019) but also for navigating the digital landscape in general (Carmi et al., 2020). Despite its evident significance, a substantial portion of the population still lacks adequate data literacy, relegating them to passive "data subjects" rather than empowered data users (D'Ignazio, 2022). This data literacy divide perpetuates inequalities and denies individuals the

Responsible Editor: Alfred Zimmermann

✉ Carolin Stein
stein@fzi.de

Timm Teubner
teubner@tu-berlin.de

Stefan Morana
stefan.morana@uni-saarland.de

¹ FZI Research Center for Information Technology, Haid-Und-Neu-Straße 10-14, 76131 Karlsruhe, Germany

² KIT Karlsruhe Institute of Technology, Kaiserstraße 89-93, 76133 Karlsruhe, Germany

³ TUB Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

⁴ Einstein Center Digital Future, Wilhelmstraße 67, 10117 Berlin, Germany

⁵ Saarland University, Campus C3 1, 66123 Saarbrücken, Germany

agency to benefit from the data-driven landscape (D'Ignazio, 2022). Consequently, the quest for effective countermeasures becomes a critical pursuit. A promising use case in this regard is citizen science (CS) (Twidale et al., 2013), referring to the participation of non-professionals in scientific research activities (Shirk et al., 2012). Historically, CS strived to democratize science and counteract social inequity (Irwin, 1995). Although participatory activities vary, in many CS projects, participants can access and work with scientific data (National Academies of Sciences & Medicine, 2018). It is hence a natural fit when thinking about conveying data literacy and shifting modes of power and agency. Currently, however, several barriers prevent the full realization of educational benefits. First, while many CS projects enable participation in data collection, participation in consecutive exploration and interpretation of data is sparse (Monzón Alvarado et al., 2020). The complexity or confidentiality of data and tasks can reduce the offer of activities (Kloetzer et al., 2021). Second, researchers and project initiators are limited in resources, such as funding and time (Kloetzer et al., 2021; Wald et al., 2016) necessary to organize participation and support. This makes current educational CS tools such as (peer) mentoring, tutorials and trainings, or curriculums potentially unsuitable as they imply additional efforts for researchers or the community (National Academies of Sciences & Medicine, 2018). While advisory bodies call for data literacy to be actively addressed in CS projects (National Academies of Sciences & Medicine, 2018), given the current challenges, appropriate solutions must first be explored. Specifically, to succeed on a larger scale, a flexible yet automated support is required. Therefore, we propose that conversational agents (CAs) might be suitable tools for this task. CAs enable the provision of support and information cost-effectively (Kvale et al., 2021) and are used in many educational settings (Okonkwo & Ade-Ibijola, 2021). They can increase learners' motivation and enable students to access content or receive help swiftly (Okonkwo & Ade-Ibijola, 2021). As a support tool for data exploration, a CA could enable citizens to participate in this research step without producing additional work or costs for initiators, such as personal training or mentoring. Also, it would be a more scalable and constantly available solution, quickly providing citizen scientists with information and assistance to enable their participation and learning. Nevertheless, compared to tutorials and curriculums, CAs can provide personalized support that can adapt to the needs of the individual citizen.

CAs require conscious design to fit the audience's and the domain's idiosyncratic requirements. Current research on CA design and utilization encompasses aspects of education (e.g., Okonkwo & Ade-Ibijola, 2021; Pérez et al., 2020), working with data (e.g., Alaaeldin et al., 2021; Narechania et al., 2021), and CS (e.g., Holowka et al., 2021; Tavanapour

et al., 2019). However, the intersection of these three topics remains a research gap. In particular, while CAs for collecting (e.g., Holowka et al., 2021; Lia et al., 2023; Tavanapour et al., 2019) and accessing data (e.g., Narechania et al., 2021; Neumaier et al., 2017; Simud et al., 2020) have received some scholarly attention, the consecutive use case of analyzing data (i.e., an integral part to strengthen data literacy) has not been explored yet. We hence seek to answer the following research question:

RQ: How should a conversational agent be designed to support data exploration in citizen science applications?

We address the research question by applying the design science research (DSR) approach. Beyond conveying data literacy, we identify the need for motivation and empowerment of citizens in the literature. Based on this, we derive design principles for a CA supporting citizen participation in data exploration and implement them in a prototypical artifact. Evaluating the prototype in an experimental study, we find that using the CA can enhance data literacy and analysis performance among inexperienced users. With this research, we aim to contribute to the ongoing efforts in reducing information disparities and ensuring that data is leveraged for societal benefit. We further identify opportunities for future research by examining the limitations and challenges of the artifact and our research approach.

Related work

As the foundation of our work, we review the literature on data literacy and its relationship to CS. We examine existing efforts to support civic engagement in data analysis and explore the potential of CAs in the domains of data literacy, education, and CS to guide our CA design.

Data literacy

Data literacy can be referred to as “the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application, and resulting value” (Panetta, 2021, para. 3). Rooting back to the notion of *information literacy*, with an increasing amount of data and the emergence of more and more data-driven professions, data literacy emerged as a buzzword in research and popular press (Schüller et al., 2019). While Gartner's definition of data literacy focuses primarily on describing a skill set (Panetta, 2021), other definitions also emphasize the ability and motivation to use these skills in one's environment. Bhargava et al. (2015) define it as “the desire and ability to constructively engage in society through and about data” (p. 24), and Schüller et al. (2019) describe it as an ability needed to navigate the digitalized

world and make informed decisions. Therefore, effective data literacy promotion should not only focus on skills but also on empowering and motivating learners to apply these skills in their respective contexts (Bhargava et al., 2015). To guide teaching and evaluation approaches to data literacy, Schüller et al. (2019) developed a data literacy framework that can be tailored to different needs and requirements. They subdivide data literacy into six core competencies: (1) the establishment of a data culture, (2) the provision of data, (3) the exploitation of data, (4) result interpretation, (5) interpretation of data, and (6) the derivation of actions (Schüller et al., 2019). While the framework provides information about the content required to promote data literacy, it does not address how it can be taught. Examples of teaching approaches to data literacy comprise in-person formats such as workshops (e.g., D'Ignazio, 2022; Debruyne et al., 2021), school initiatives (e.g., Bhargava et al., 2016; Gould, 2021), or online formats such as forums, quizzes, and online classes (Jayawickrama et al., 2020). In addition, many digital tools facilitate data-related tasks, such as data collection, processing, and visualization. D'Ignazio and Bhargava (2016) have mapped tools such as Excel, cartoDB, or infogr.am in view of their flexibility and expertise requirements. However, they pointed out that current tools emphasize output creation rather than learning. They derive four design principles for pedagogical learning tools: targeted focus, guidance, inviting design, and tool expandability (D'Ignazio & Bhargava, 2016). These principles should ensure that tools ease barriers to learning and quickly get users started with activities. While being invited to follow appealing first activities, users should find additional information on more demanding practices (D'Ignazio & Bhargava, 2016). Other authors stress that teaching approaches should encompass multiple pathways for users to choose from—according to their needs (e.g., Bhargava et al., 2015). These should be agile, adaptive, and focus on what is effective and meaningful for the learners, such as working with community data (Bhargava et al., 2015; D'Ignazio, 2022).

Citizen science

Citizen Science is defined as “the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources” (Socientize, 2014, p. 6). Originally used in the natural sciences, today CS has proven useful across many different fields (Pettibone et al., 2017). With the expansion of CS, the heterogeneity of participation approaches has also increased (Shirk et al., 2012; Spasiano et al., 2021). The most common project types are contributory projects that focus on participatory data collection (Bowser et al., 2020; Monzón Alvarado et al., 2020). The participatory analysis and interpretation of data is less common and usually occurs in co-created or collegial CS

projects (Shirk et al., 2012). However, citizens have increased access to raw data, for instance, through open (government) data platforms. They could support public institutions in drawing important insights, when given access to (gamified) toolkits supporting data utilization (Krishnamurthy & Awazu, 2016; Simonofski et al., 2022; Wirtz et al., 2022). On the CS platform *Zooniverse* (www.zooniverse.org), for instance, a strong focus in data analysis is put on participatory image classification (Bonney et al., 2016; Simpson et al., 2014). Moreover, for individual CS projects, digital tools such as Google Spreadsheet are prepared but often used only in a classroom setting, where additional support and teaching are provided (e.g., Kjølvik & Schultheis, 2019; Shah & Martinez, 2016). Within CS projects, learning happens either on a micro level (e.g., through active participation and the execution of tasks) or on a macro level (e.g., by sharing videos or online tutorials) (Jennett et al., 2016). However, the education of citizen scientists and, thus, the provision of educational tools are rarely the focus of CS projects. In a study exploring participant motivation and retention in digital CS projects, Wald et al. (2016) reported that for most projects, scientific outcomes were the focus while “educational and social benefits [...] were incidental” (p. 562). For researchers, the main barriers to learning are the necessary temporal, technical, or monetary resources, as well as breaking down complex tasks (Kloetzer et al., 2021). Conversely, participants can be prevented from learning due to a lack of confidence, skills, money, or time (Kloetzer et al., 2021). In addition, project design itself can negatively influence learning by including too little feedback or interaction (Kloetzer et al., 2021). These obstacles should be a starting point for technical solutions supporting participatory data exploration.

Conversational agents

CAs are applications that allow users to interact with them in a natural language and can either be text or speech-based (Janssen et al., 2020; Rapp et al., 2021). They are also discussed in certain domains under the terms chatbot or chatterbot (Bittner et al., 2019). Offering automation where priority human resources are needed, CAs can be a low-threshold solution, leading to large cost reductions (Kvale et al., 2021). However, maintaining user satisfaction can pose major challenges. Studies on customer service chatbots indicate that factors such as problem resolution, answer precision, and concreteness drive customer satisfaction (e.g., Kvale et al., 2021; van der Goot et al., 2021), while errors and a lack of functionality can quickly deteriorate it (e.g., van der Goot et al., 2021). Likewise, studies on CAs in the workplace suggest that CA adoption depends on user characteristics such as individual tech savviness (e.g., Gkinko & Elbanna, 2023). The possible application fields for CAs range from economics (e.g., finance or e-commerce) to personal applications

such as health or emotional support (Rapp et al., 2021). It has been shown that good CA design highly depends on the domain it is built for (e.g., Bittner et al., 2019). Further research on the transferability of design knowledge between contexts is necessary (Diederich et al., 2022). For instance, while the usage of social cues is encouraged in some CA applications (e.g., Holowka et al., 2021; Tavanapour et al., 2019), it can have detrimental effects when the reliability of information is essential (Stieglitz et al., 2022). Thus, a one-size-fits-all approach to CA design is unrealistic—context, stakeholders, and unique value propositions must be considered (Janssen et al., 2020). To guide the design of a CA to support citizens in data exploration, different domains provide interesting insights. In the following, we shed light on insights from the application of CAs for education, (big) data-related work and CS (see also Table 1 in the electronic supplementary material).

In the domain of education, research differentiates between teaching- and service-oriented CAs. While the former describes CAs targeting knowledge generation, service-oriented CAs provide administrative services, such as introductory or library services (Pérez et al., 2020). When interacting with learners, CAs usually act as “teacher, student, or colleague” (Tamayo-Moreno & Pérez-Marín, 2016, p. 1). In this role, CAs have proven beneficial as they allow for integrating multiple content into one tool and parallel access by multiple users (Okonkwo & Ade-Ibijola, 2021). The possibility of receiving immediate help on demand is convenient with positive effects on learning motivation (Okonkwo & Ade-Ibijola, 2021). CAs also proved suitable for closing learning gaps between mainstream learners and learners from certain minority groups (Pérez et al., 2020). However, through a structured literature review, Pérez et al. (2020) have identified boredom and user frustration (e.g., through lengthy messages and inadequate replies) as common impediments. Teaching CAs can target various topics and domains, with language learning being a prominent use case (Pérez et al., 2020). Another use case, closer related to data literacy, is math education where CAs have been used (e.g., Anh & Ngan, 2021; Nguyen et al., 2019).

CAs in data-related work environments usually focus on data provision and depiction for non-technical skilled employees (e.g., Alaaeldin et al., 2021; Narechania et al., 2021; Simud et al., 2020). CAs can conduct tasks such as generating database queries and visualizations in or based on natural language (e.g., Hoon et al., 2020; Narechania et al., 2021; Neumaier et al., 2017; Simud et al., 2020). They can also support decision-making by explaining analytic tools and key performance indicators for a given dataset (e.g., Alaaeldin et al., 2021). Another important application is the identification of relevant datasets, domain-specific scientific tools, and methods (e.g., Keyner et al., 2019; Neumaier et al., 2017; Zhang et al., 2019). Overall, CAs in data-related

work environments mainly focus on overcoming modern databases’ technical complexities through natural language interfaces. Since supporting the understanding of data and analysis methods is not the focus of publications, fundamental data literacy remains a prerequisite for users.

Within the domain of CS, the usage of CAs for different activities is not yet a common practice. First and foremost, CAs have been used for quantitative and qualitative data collection in CS projects (e.g., Holowka et al., 2021; Isacco et al., 2018; Lia et al., 2023; Tallyn et al., 2018; Tavanapour et al., 2019). They enable participants to answer questionnaires, upload text, pictures, or geotags (e.g., Isacco et al., 2018; Lia et al., 2023; Tallyn et al., 2018) and can, in return, provide guidance or encouragement and support or share data directly with experts or the community (e.g., Holowka et al., 2021). Advantages of their use in data collection can include personalized feedback and the ability to conduct further inquiries when observations are incomplete (Portela, 2021). Additionally, they can provide citizens with data or visualizations (Portela, 2021). Other work explores the advantages of CAs facilitating the ideation process by collecting, structuring, and presenting ideas (e.g., Tavanapour et al., 2019) or supporting the community and its interaction (e.g., Athreya et al., 2018; Portela, 2021). Overall, the potential of using CAs for CS seems to be not yet exploited. For example, we could not identify literature presenting a CA used for training citizen scientists, although the suitability of CAs for educational purposes has been proven in other domains (Okonkwo & Ade-Ibijola, 2021).

Research gap

The related literature on data literacy, CS, and CAs offers valuable insights into the possibilities and challenges associated with designing activities and tools for enhancing data literacy. However, it underscores a significant research gap at the intersection of these topics: the design of tools for active participation in data exploration. While the data literacy literature provides crucial insights into tool design for learners (e.g., D’Ignazio, 2022; Schüller et al., 2019), it emphasizes the need for more learning-oriented tools embedded in a meaningful context for the user (Bhargava et al., 2015; D’Ignazio & Bhargava, 2016). The CS literature addresses this context and discusses participants’ learning and tools to support projects (e.g., Jennett et al., 2016; Liu et al., 2021; National Academies of Sciences & Medicine, 2018). However, it reveals numerous challenges in integrating educational components (e.g., Kloetzer et al., 2021; Wald et al., 2016) and that the data analysis step is often not addressed. The educational CA literature generally discusses opportunities and challenges in using CAs for teaching (e.g., Okonkwo & Ade-Ibijola, 2021; Pérez et al., 2020), but in the specific context of working with data, the focus remains

primarily on discovering datasets (e.g., Keyner et al., 2019; Neumaier et al., 2017; Zhang et al., 2019) or making them accessible (e.g., Hoon et al., 2020; Narechania et al., 2021; Neumaier et al., 2017). Likewise, CAs in CS do not focus on facilitating data analysis but rather center on qualitative or quantitative data collection (e.g., Holowka et al., 2021; Lia et al., 2023; Tallyn et al., 2018). Additionally, they do not exploit their teaching capabilities.

Overall, the existing literature provides crucial insights for designing CAs supporting citizens in data exploration, yet specific design guidelines for this use case are missing. Considering the challenges on the transferability of design knowledge across contexts (Bittner et al., 2019; Diederich et al., 2022; Janssen et al., 2020), this represents a significant research gap that should be further explored and investigated.

Research approach

To answer our research question, we conducted a research project following the DSR approach of Peffers et al. (2020). For a “problem-centered approach” (Peffers et al., 2020, p. 56), the methodology includes six steps, starting with identifying a problem and deriving a motivation for its solution (see Fig. 1). First, to elaborate on the problem to solve, we assessed the current state of research by reviewing related literature in the field of data literacy, CS, and CAs and performed a stakeholder analysis based on insights from these domains. In addition, we carried out an expert workshop on data analysis conduction with 12 experts and advanced practitioners, combining ideas from two established requirements elicitation methods “Introspection” and “Brainstorming” (Sharma & Pandey, 2013). While in an introspection, experts elicit the user needs based on their domain knowledge; in brainstorming, participants from different stakeholder groups are invited to collectively generate

ideas (Paetsch et al., 2003). The expert workshop facilitated introspection through a think-aloud session about “conducting a data analysis.” Think-aloud sessions, typically known from usability testing, enable researchers to gain insights into participants’ thought processes by neutrally observing them speaking out their thoughts aloud while working on a given task (Ericsson & Simon, 1984; Fan et al., 2020). After the individual think-aloud session, brainstorming was conducted in the expert workshop as an open discussion between all participants, guided by the components of the data literacy framework (Schüller et al., 2019). The approach to the conduct of the expert workshop is further described in the “[Expert workshop \(activity 1\)](#)” section. Second, the following activity of the DSR approach comprised the definition of solution objectives by deriving either quantitative or qualitative requirements for the artifact (Peffers et al., 2020). Therefore, we translated the results from the first activity into atomic user needs as solution objectives and positioned them in the related literature. Thirdly, we approached the actual design and development phase by instantiating our artifact. The two tasks of this phase were outlining the necessary function and design requirements before practically creating the artifact (Peffers et al., 2020). Using our user needs, we determined design specifications, following the schema by Gregor et al. (2020), and implemented them in a prototypical instantiation. The approach to specifying the design principles and implementing the artifact is further described in the “[Design principles for a conversational agent for public participation in data analysis](#)” and “[Artifact](#)” sections, respectively. To close the first DSR cycle, the artifact should be demonstrated and evaluated. We covered these steps simultaneously by designing and conducting a final experiment with 30 participants guided by best practices to evaluate CAs and data literacy learning (Pérez et al., 2020; Schüller et al., 2019) and DSR artifacts (Venable et al., 2016). Using a between-subjects design, the experiment compared the data

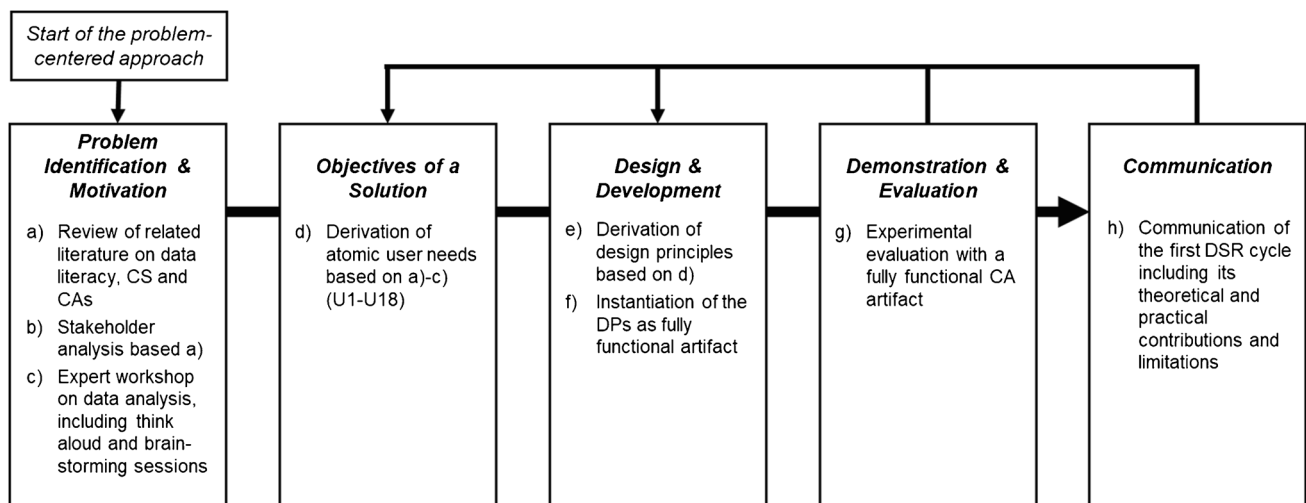


Fig. 1 Overview of the DSR approach

analysis performances and (learning) experiences of a participant group guided by the designed artifact with the results of a group that received no explicit guidance but was allowed to use existing material through the web. By combining quantitative (task performance, empowerment, motivation, perceived learning, user experience) and qualitative (user experience, dialog analysis) insights, the experiment enabled us to evaluate the artifact comprehensively. The approach to the experiment's design, conduction, and evaluation is described in detail in the "Evaluation" section.

Designing a conversational agent for public participation in data analysis

Problem awareness and solution objectives

We set out to design a CA capable of supporting data exploration in CS projects. This design endeavor entails understanding the intricacies of data literacy and CS and applying this knowledge to CAs. The review of relevant literature has shown that fostering data literacy requires creative teaching approaches that consider the interests and realities of the target audience

(e.g., Bhargava et al., 2015; D'Ignazio, 2022). A challenge in this regard is the complexity of the above concepts and that data literacy is based on a set of competencies rather than a specific skill or technique (Debruyne et al., 2021; Schüller et al., 2019). The intended data literacy content thus needs to be broken down to guide the design of the CA (activity 1).

While data collection is frequently seen in CS projects (Bowser et al., 2020; Liu et al., 2021; Monzón Alvarado et al., 2020), few projects use participatory data analysis (beyond classification tasks). In general, budgets and time constraints limit such projects for researchers and citizens (Kloetzer et al., 2021; Wald et al., 2016). In addition, maintaining momentum and keeping citizens engaged and active (beyond the first exploration) are crucial but difficult (Wald et al., 2016). Therefore, understanding citizen scientists is key to the design of appropriate support tools (activity 2).

Expert workshop (activity 1)

To understand the specific requirements of the CA (content and design), we invited 12 data analysts first to an individual think-aloud session and second to a brainstorming session. The participants consisted of six Ph.D. candidates and six

Table 1 Atomic user needs grouped by perspective and contrasted with related literature

User needs derived from the workshop		Concepts in the literature
As a student , I want to...	U1 Understand the data analysis process	Different activities in data value creation (Schüller et al., 2019)
	U2 Understand the dataset (meaning, usefulness)	Competence <i>obtain data, prepare data</i> (Schüller et al., 2019)
	U3 Know and apply methods of data cleaning	
	U4 Find a start for the data analysis	Low entry point for data literacy tools (D'Ignazio & Bhargava, 2016)
	U5 Understand and select a method for data analysis	Competence <i>analyze data, interpret data analysis, visualize data, interpret data visualizations</i> (Schüller et al., 2019)
	U6 Note limits and challenges of the data analysis	
	U7 Interpret analysis results and critically question findings	
	U8 Select and design appropriate data visualizations	
	U9 Understand how to interpret and check visualizations	
	U10 Ask questions	CA's answering student's questions (Okonkwo & Ade-Ibijola, 2021)
	U11 Decide on an analysis action and path	Enabling multiple pathways for learners to choose from (Bhargava et al., 2015)
	U12 Get access to data analytics tools	Knowledge and mastery of tools as essential skills (Schüller et al., 2019)
	U13 Get access to assistance and helpful material	CA's integrating multiple contents (Okonkwo & Ade-Ibijola, 2021)
As a tutor , I want to...	U14 Steer through the analysis process	Guided data literacy tools (D'Ignazio & Bhargava, 2016)
	U15 Point out missing competencies and skills	-
	U16 Proceed with questions (answer, redirect, collect)	CA's answering student's questions (Okonkwo & Ade-Ibijola, 2021)
	U17 Form an interface to other material/ software	Expandable data literacy tools (D'Ignazio & Bhargava, 2016)
	U18 Support different levels of pre-knowledge	CAs providing individualized support (Okonkwo & Ade-Ibijola, 2021)

master students majoring in data science-related fields and presented thus advanced to professional practitioners in the field. Sessions were conducted virtually using an online conference tool moderated by one researcher, which had some implications for the workshop conduction. In the think-aloud session, participants were asked to solve several analytical tasks based on a given dataset. Instead of being able to observe their actions physically, in the virtual think-aloud sessions, we used audio, video, and screen sharing to enable as many insights as possible for the researcher. In the following group brainstorming session, participants could then discuss approaches and pitfalls to data exploration and requirements for support based on their own experiences in the think-aloud session. Since in virtual compared to on-site groups lower social presence can be a challenge to the discussion quality, we limited the size of the brainstorming groups to three participants per session to better integrate the individual participants and utilized an online whiteboard to facilitate collaboration (Roberts et al., 2006). The results of the expert workshop were translated into atomic user needs of students in data exploration and tutors in student support and contextualized with the literature review results (see Table 1).

Stakeholder analysis (activity 2)

Most CS initiatives cater to broad audiences (Spiers et al., 2019). Common user characteristics include above-average education, above-average income, and above-average seniority (Ciarán Mac Domhnaill & Nolan, 2020; National Academies of Sciences & Medicine, 2018). In addition, citizen scientists tend to “embody the characteristics of autonomy, competence, and relatedness in their hobby” (Jones et al., 2018, p. 15). However, CS embraces the diversity of participants, and project organizers claim to strive for more diversity in terms of age, gender, and ethnicity (National Academies of Sciences & Medicine, 2018). Thus, the level of autonomy and knowledge can be assumed to be heterogeneous. Based on this and research on data literacy (e.g.,

Logan, 2017; Watson & Callingham, 2004), we distinguish the following (broad) user groups to have in mind for the CA artifact: (1) beginner users, (2) advanced users, and (3) professional users (see Table 2).

Design principles for a conversational agent for public participation in data analysis

A design principle should include an aim, implementer, and user; a context; mechanisms; and a rationale (Gregor et al., 2020, p. 1634). We follow this scheme and propose five principles for the design of a CA for support in data analysis tasks (context), which can be used by researchers and developers (implementer) to create software support for non-expert citizen scientists undertaking their analytical activities (users). Considering user needs U1, U11, U14, and U18, the platform must specify a certain process structure while the user remains free to choose how to follow this path. We find guidance for this requirement in the design principles of Tavanapour et al. (2019), who state that a CA for idea creation must be able to follow a given conversation flow while still being able to lead the process actively and D’Ignazio and Bhargava (2016), who underlined, a data literacy tool should be guided. Additionally, Portela (2021) advises that a CA should include fixed chat commands for user orientation. Therefore, we formulate the first design principle as follows:

DP1: In order to structure the analysis process (aim), the system should provide a sorted menu highlighting the individual parts of a data analysis (mechanism), as this enables the user to get guidance on the process and navigate to a specific topic of interest (rationale).

Several user needs (U2, U4) express that beginners’ entrance must be eased. Thus, the system should “provide a low entry point” (p. 87) for data analysis (D’Ignazio & Bhargava, 2016). Nevertheless, the knowledge needed to conduct many such tasks is comprehensive (U3, U5, U7, U8, U9). The stakeholder analysis showed the need to account for different user groups, which is supported by Bhargava et al.

Table 2 Description of user groups for the CA artifact

User groups	Description
Beginner users	No to little knowledge about data literacy (see idiosyncratic level, informal level (Watson & Callingham, 2004)) New to CS or without the typical characteristics of an experienced citizen scientist, requiring guidance and explanations
Advanced users	Experienced citizen scientists or people experienced in data literacy with a fundamental understanding of the data process (see conversational level; Logan, 2017), data analysis, and visualization methods (see inconsistent and consistent non-critical level; Watson & Callingham, 2004) Knowledge is rather basic and incomplete or might date back a long time ago, but users are more autonomous and inform themselves or might have specific questions
Professional users	Familiar with working with data (see critical and critical mathematical level; Watson & Callingham, 2004) and working largely to completely autonomously Group is rather out of scope for the CA

(2015), pointing out the necessity to provide “multiple pathways for people with different data literacy needs and capacities to interact within a complex system” (p. 15). Therefore, the platform should equip users with the appropriate background knowledge based on their needs and interests. Tavanapour et al. (2019) to this end propose a comparable mechanism specifying that CAs should have the “capacity to summarize [...] information [...] and offer further explanations, if requested” (p. 8). We, therefore, formulate our second design principle as follows:

DP2: The system should provide a tiered knowledge structure (mechanism) to educate the user efficiently (aim), as this enables the user to determine the depth according to their interests and skills (rationale).

User needs U10 and U16 express that the users should be enabled to get answers to their specific questions, which is a common functionality for teaching CAs (Okonkwo & Ade-Ibijola, 2021). We, thus, formulate the third design principle as follows:

DP3: The system should allow users to enter questions and process them (mechanism) to get answers to individual questions (aim, rationale).

User needs U12, U13, and U17 imply that the platform should use existing teaching material. To this end, D’Ignazio and Bhargava (2016) point out that data literacy tools should be expandable, bridging the pathway for learners to go from one data literacy tool to the other. We incorporate these findings in the fourth design principle:

DP4: To efficiently educate the user (aim), the system should provide a combination of self-developed and external materials through embedding or forwarding (mechanism), as users have an interest in a broad offer of learning material (rationale).

Furthermore, the presence of many pitfalls (user needs U6, U9, U15) requires the platform to support users in understanding challenges and avoiding common mistakes. We, therefore, propose that:

DP5: The system should provide indications and warnings of challenges and common mistakes in time (mechanism) to prevent the user from failing (aim, rationale).

An overview of the design principles and their derivation from the atomic user needs are found in Table 3.

Artifact

In the third phase of the DSR process, the formulated design principles are instantiated in an artifact in the form of a CA prototype. The CA provides dataset-independent support to beginner and advanced users in the process of data analysis. It provides process-oriented and knowledge-based advice through messages, pictures, links, and guidance along two workflows:

Workflow 1

The data analysis (DA) workflow offers guidance for beginners and provides a menu of steps (Fig. 2), showing different steps of a typical data analysis process (DP1). The menu serves as a central point to which the user returns within the flow. The first step of the DA process (“Getting started”) reflects DP2, DP4, and DP5. After receiving information on how to get started, the user can request more information (DP2) or browse through external education material (DP4). To address DP5, the CA invites users to analyze their data actively. Upon the user’s confirmation to proceed, the bot provides an overview of common mistakes concerning the task the user has just completed (DP5).

Workflow 2

The question and answer (Q&A) workflow should attract users with basic data knowledge. Here, users can specify topics of interest by asking questions. Upon a request, the CA either recognizes the question as dataset-specific or methodological. In the former case, the bot points out that such questions are out of scope. In the latter case, the bot provides an answer if it recognizes the question. If the question is not recognized, the CA offers to forward the question to a supervising researcher. Upon affirmation, the CA sends questions and contact details to an online spreadsheet, privately accessible to the supervising researcher.

Table 3 Design principles mapped to their respective user needs

Design principles	Associated user needs
DP1 Provide a menu that structures the analytics process for the user to get guidance or navigate to a specific topic	U1, U11, U14, U18
DP2 Provide a tiered knowledge structure to let the user determine the depth according to their interests and skills	U2, U3, U4, U5, U7, U8, U9
DP3 Allow the users to enter questions and process them by either ad hoc answering or forwarding	U10, U16
DP4 Provide a combination of self-developed and external materials through embedding or forwarding to educate the user efficiently	U12, U13, U17
DP5 Provide indications and warnings of challenges and common mistakes in time	U6, U9, U15

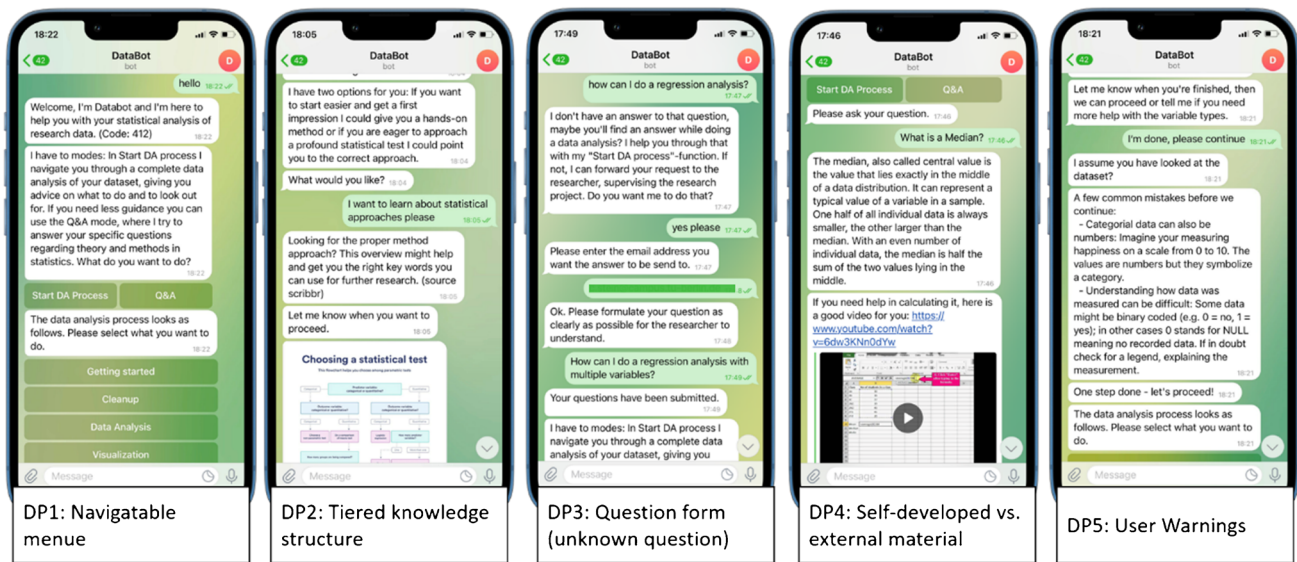


Fig. 2 Exemplary conversations with the CA reflecting the implementation of the design principles

The implementation of our CA is based on the open-source programming framework *Rasa Open Source Machine Learning Framework*. We used a local setup with the default NLU and core component configuration for the prototype. In addition, Rasa X has been used to facilitate conversation-driven development, repeatedly asking potential users to test the CA in its different stages of development. To lower the barrier of usage, we chose to implement the front end via Telegram (i.e., a popular messenger service).

Evaluation

The evaluation phase of a DSR project assesses whether (and if so, how) the artifact solves the problem (Peffer et al., 2020). While for customer service bots it is often sufficient to determine the share of adequate responses, the degree of success in education applications depends on the learning effect generated for the user (Pérez et al., 2020). Thus, the evaluation can, for instance, rely on the learner’s perception measured through questionnaires or on comparison with a control group, not utilizing the CA (Pérez et al., 2020). For the evaluation of data literacy, Schüller et al. (2019) propose the *stage model* by Kirkpatrick, which assesses enjoyment, learning success, the learner’s behavior, and learning outcomes (Kirkpatrick, 1959). Naturally, behavior and (long-term) learning outcomes can only be evaluated to a limited extent, especially when drawing on online experimental methods. We thus combine approaches from the CA and data literacy perspective in a between-subjects experiment. The experiment consists of an initial questionnaire, the main experimental part, and a post-questionnaire. The initial questionnaire (Table A2 in the electronic supplementary

material) is used to evaluate prior knowledge of the study participants, such as knowledge about CS, data analysis, and experience in working with datasets. Additionally, it assesses their motivation for science, including aspects of intrinsic and career motivation, and self-efficacy, using items from the Science Motivation Questionnaire (SMQ), a well-established instrument in pedagogic work and research (Glynn et al., 2011). In the practical part, participants are introduced to the topic of CS and the experiment task, which is a data exploration of the data of different formats on passengers on the Titanic (e.g., age (numerical), embarkation port (categorical), survival (binary)). This open-source dataset is free of charge and is known for its use in introductory courses, as well as academic studies (e.g., Ekinici et al., 2018; Gupta et al., 2018), making it particularly suitable for the experiment. The participants receive an extract of the dataset in the form of a .csv and .xlsx file and are asked to complete 12 practical and theoretical data exploration tasks (Table A1 in the electronic supplementary material). Following a between-subjects design, one user group may use our CA artifact for these tasks. They obtain an introduction to the CA and its functionalities and are asked to install it on their device. In contrast, the control group does not receive the artifact. To depict the current status quo, this control group is advised that it is allowed to use all other existing software or learning material, for instance, through the web. The overall task performance of participants is calculated based on all 12 tasks and normalized to the interval [0, 1]. The final evaluation quantitatively assesses user motivation, empowerment, and perceived learning, providing a contrasting introspective view of task performance. To do so, we make use of established survey constructs, i.e.,

interest/enjoyment from the Intrinsic Motivation Questionnaire (Center for Self-Determination Theory, 2022; Ryan, 1982), a second-order construct for empowerment (Kim & Gupta, 2014), and a construct for perceived learning (Alavi et al., 2002) measured on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7). In addition, for the CA treatment, user experience is assessed via the User Experience Questionnaire (UEQ) (Schrepp et al., 2017). To get qualitative input, we use further open-ended questions and analyze participants' conversations with the CA. The post-questionnaire is found in Table A3 of the electronic supplementary material.

Testing

First, we conducted a set of pretest sessions, in which five users tested the CA treatment and one user tested the control treatment.

Procedure

The experiment was executed between December 2021 and January 2022. Participants were recruited through the online platform Prolific (e.g., Palan & Schitter, 2018). Although the experiment was conducted asynchronously, participants could contact a supervising researcher in case of any issues or technical difficulties during the sessions. The average payout was 6.21 GBP per hour, and the average completion time was 46 min. From those, participants spent 39 min on average processing the analytical tasks.

Sample

The sample included $n = 30$ international participants who self-reported fluency in English. Participants were between 19 and 45 years old (23 male, 7 female) and could mostly be allocated to the level of beginner to advanced users: 37.67% of the participants had never worked with datasets before, while the average pre-knowledge of data analysis was indicated with 4.23 points (see Table 4). Only one participant indicated a very high level of data analysis pre-knowledge

potentially presenting a professional user. The sample was split into $n_{\text{control}} = 10$ participants for the control treatment and $n_{\text{CA}} = 20$ for the CA treatment. This distribution was chosen for several reasons. Firstly, the primary focus of the study was to evaluate the design and effectiveness of our CA by assessing its usage through real users. Thus, by allocating a larger sample to the CA treatment, we collected more comprehensive and diverse insights into use behaviors and perceptions. Secondly, while the experiment design enables us to evaluate user interaction with the CA, interaction with self-selected materials and sources in the control treatment cannot be tracked. Therefore, uncovering different types of users or approaches by opting for a larger sample size was only feasible for the CA treatment. The importance of the control group was to evaluate the relative effectiveness of the CA. Overall, we opted for a smaller sample size in favor of longer experiment duration to ensure that participants were sufficiently engaged with the tasks at hand.

Randomization check

To ensure treatment randomization, the distribution of age, gender, and interest- and knowledge-related factors were evaluated. There were no significant age differences between the two treatment groups, with a mean age of 27.1 years in the CA treatment and 27.5 years in the control treatment ($p = 0.558$). Similarly, the gender distribution was not significantly different, with 20% and 30% female participants in the CA and control treatment, respectively ($p = 0.885$). The two-sided t -test results for the remaining factors involving pre-knowledge and interest also showed no significant differences between the mean scores of the two treatments (Table 4). Therefore, we assume treatment randomization was successful.

Quantitative results

We now assess how the availability of the CA affected participants' perceptions of task performance (TP), perceived learning (PL), empowerment (Emp), and motivation (Mot). The data is summarized in Table 5 and shown in Fig. 3.

Table 4 Randomization checks for interests and pre-knowledge of experiment participants

Treatment		Data analysis knowledge	Software skills	Dataset experience*	CS concept*	SMQ
CA	Mean	4.400	5.700	0.650	0.100	6.058
	SD	1.536	0.923	-	-	0.694
Control	Mean	3.900	4.900	0.600	0.200	5.666
	SD	1.792	1.287	-	-	0.568
p -value		0.433	0.060	0.797	0.465	0.134

Variables have been measured on a 7-point Likert scale, except for *, which were binary variables. p -values refer to the results of two-sided t -tests

Table 5 Summary statistics TP, PL, Emp, and Mot by treatment

Treatment		TP	PL	Emp	Mot
CA	Mean	0.630	5.640	4.979	5.113
	SD	0.167	1.077	0.925	1.490
Control	Mean	0.488	5.960	4.925	5.775
	SD	0.197	0.799	1.102	0.786
<i>p</i> -value		0.048	0.414	0.888	0.122

p-values refer to two-sided *t*-tests or Welch tests, respectively

Task performance Task performance was measured on a scale from 0 to 1, where 0 indicates no task has been solved correctly, while 1 indicates a participant has solved all tasks correctly. We find that task performance is significantly higher in the CA treatment than in the control condition ($p=0.048$), with a relative surplus of 14%. Beyond overall task performance, we also assess each task individually. An overview of the results can be found in the appendix. Notably, for all but two tasks, participants from the CA condition performed better than their counterparts from the control condition.

Perceived learning, empowerment, and motivation Overall, participants indicated substantial perceptions of learning (i.e., 5.75 points on the 1–7 Likert scale), empowerment (i.e., 4.96 points on the 1–7 Likert scale), and motivation (i.e., 5.33 points on the 1–7 Likert scale). For all variables, however, we do not find significant differences between treatment conditions (see Table 5).

Impact of age, gender, and pre-knowledge We do not find any effects of age or gender on any of the target variables. For participants’ pre-knowledge, we see a small effect on task performance ($\beta=0.04$, $p=0.047$). In addition, we observe small correlations between SMQ ($p=0.005$) and software skills ($p=0.030$) with perceived empowerment.

User experience Overall, users sent between five and 44 messages to the CA (mean, 16.35; SD, 9.670; total, 327). During their interaction with the CA, they followed different paths: Most users (55%) initially followed the DA analysis path in the proposed order. However, some users (25%) used the DA menu to jump to the topics of interest directly. Moreover, 20% of users used the Q&A function rather than the DA Process path.

The different user pathways and transitions are illustrated in Fig. 4. On average, the CA made between 0 and 3 false intent classifications (mistake) within the user conversations (mean, 0.8; SD, 0.834). Most often, they appeared in the visualizations part of the DA process. In addition, in three conversations, the CA could not answer a question that would have been in scope (QA other questions). The user experience during the conversations was evaluated with the short UEQ, including eight items. On average, a participant rated an item with 4.65, slightly above the center value. For hedonic items, the rating was lower, with an average score of 4.45; for practical items, the average score is 4.85. The best score could be achieved in the category constructive vs. supporting, while the lowest score was measured for boring vs. exciting.

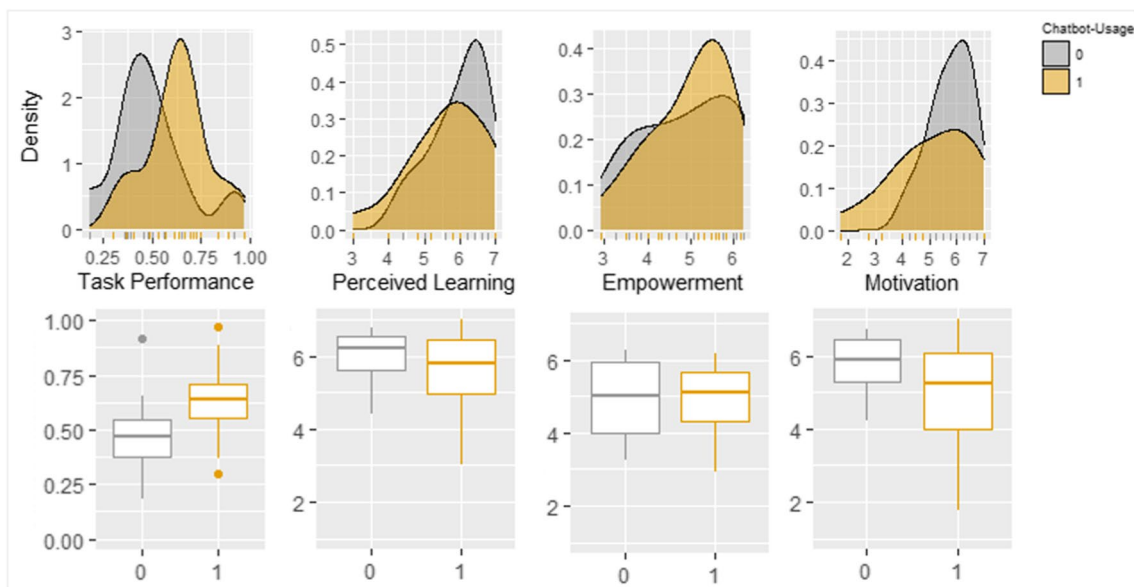
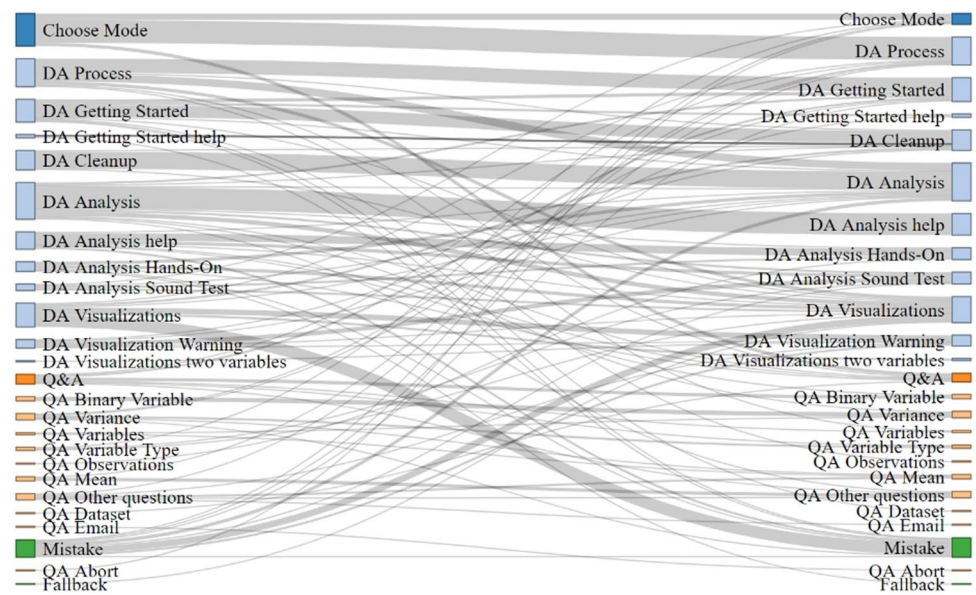


Fig. 3 Distribution and score comparison for TP, PL, Emp, and Mot grouped by treatment

Fig. 4 State transition conversations with the CA



Qualitative results

Within an open question asking for feedback about the experiment, both groups were highly satisfied and stated that they enjoyed the experience. One participant from the CA treatment stated: “I would like to participate more in studies using this kind of educational bot” (P1) while another denoted it as a “fun experience” (P28). On the other hand, a control group participant stated: “This study was for sure challenging, which I tend to enjoy [...]” (P30). Furthermore, both groups were asked about the support they used and additional support they would have liked. In the control group, the most common support tools used were the Internet (70%) and Microsoft Excel (70%), while some participants stated that they had used a calculator or had drawn on pre-existing knowledge. For the CA group, similar tools were mentioned. However, the usage of the Internet (40%) and Excel (35%) was less frequent. One participant indicated the usage of videos suggested by the CA, while multiple participants mentioned YouTube, leaving open which videos they had watched specifically. Regarding requested support, the control treatment brought forward several ideas: Most frequent proposals were process guidance, more context information, and help with calculations such as formulas or step-by-step explanations. One control group participant formulated: “I would have liked a person to guide me through the process, I did not know where to start” (P4). Another stated: “I believe a [...] better explanation on how to go [about] the analysis process for a [newbie] would have done justice” (P26). Other ideas from control group participants comprise terminology or Excel explanations and the integration of examples. In contrast, the CA group proposes concrete improvements and extensions to

the CA. Most frequently, they formulate the desire to be able to ask dataset-specific questions. Other repeatedly mentioned ideas are automated calculation support, the ability to ask more detailed questions, and to have more examples at their disposal. Some participants want supporting graphs, more buttons, and more hints. In addition, users mentioned concrete criticism of the CA’s current implementation. One participant indicates that the introductory part should be designed to be more accessible and catchy: “The data bot can be a great tool to get [youngsters interested] in stats, but is missing a proper introduction that can spark the [interest]” (P22). Two participants added that they were confused about the conversation structure, while others thought the CA was too text-heavy. A participant formulates: “I was a little confused on if I had to follow the exact steps the Bot was leading me to, or if I could ask a question totally different from what the bot was telling me” (P3). Overall, participants from the CA treatment had varying levels of satisfaction with the artifact. Positive statements indicated that it provided appropriate support and satisfactory performance. Specifically, one participant stated: “The chatbot was all I needed” (P21). Another said, “The bot was more competent than I expected. Thanks to such bots, anyone can analyze data” (P1). Critical statements indicated different reasons for dissatisfaction. Interesting quotes were, for example: “it is easier to find the necessary info on the internet” (P13), “it would have been better if someone was on a call ready to answer any concerns or difficulties I had” (P28), or “The support was ok, but a crash course or sample problem would have been better” (P30). Finally, participants from the CA treatment group were asked whether and where they could imagine the usage of the CA other than in CS projects. Overall, 19 out of 20 participants proposed several application

fields, including analysis courses, for instance, in school or university, research, or companies.

In summary, the qualitative results highlight that participants of both treatment groups generally enjoyed the experiment, utilizing different tools for data analysis and having numerous ideas for structuring further support or improving the CA. Sentiments from both treatment groups are highly relevant to evaluate the need for existing functions and identify additional requirements. Moreover, insights into the varying levels of tool satisfaction of the CA treatment group are particularly interesting for understanding user behavior and possible groups.

Discussion

Summary

This work presents the outcome of a DSR project that aimed at developing a CA for scientific data exploration with citizens. Building on literature and a qualitative study, we gathered user needs for a CA that assists data exploration. We then established five design principles and implemented them in a prototypical application. The prototype was evaluated by an online experiment and benchmarked against self-organized tools concerning perceived learning, empowerment, motivation, and actual performance in a series of analytical tasks. The experiment's findings provide insights into the CA's design and the particularities of different users, which are reported in the following.

The CA design and effectiveness

Considering DP1-DP5, the analysis of the CA's conversations revealed that participants used both the DA process flow and the Q&A flow (DP3), and their usage behaviors demonstrate that the tiered knowledge structure and the freedom to follow the process or navigate to specific topics (DP1, DP2) were effectively implemented. Notably, participants used the Q&A flow to ask questions but did not utilize the option to forward a question. This is an important aspect to consider when evaluating the effectiveness of DP3. Regarding the learning material offered, most participants acknowledge the learning support offered by the CA. Some participants also highlighted using the external material provided (DP4). Regarding DP5, the better performance of CA participants in one particular task could indicate that the respective warning issued by the CA was adequate and successful. In terms of effectiveness, overall, participants provided access to the CA performed significantly better than those using self-organized support as a control. Interestingly, this effect was not reflected in participants' self-perceptions, and no significant correlation between task performance and

perceived learning could be measured. Although educational studies report similar observations (e.g., Barzilai & Blau, 2014; Vinales et al., 2019), this issue might hold important implications. Considering established IS theories, the participants' impression of the technology's usefulness affects usage intention (Davis, 1989). In this regard, users could unjustly reject an educational tool that does not affect perceived learning, although task performance improves, pointing towards untapped potentials of these technologies. In terms of motivation and empowerment, our findings did not provide evidence of a positive impact of the CA on the participants. Possible explanations may be related to issues with the CA's design. Literature suggests that long text messages and inadequate responses can lead to user boredom and frustration (Matsuura & Ishimura, 2017; Pérez et al., 2020). We found that some users experienced these issues, potentially impairing motivation. Additionally, UEQ results showed that the practical and hedonic quality of the CA could be improved. Compared to an industry benchmark (e.g., Hinderks et al., 2018), these scores indicate that it is in an early stage of development and needs further quality refinement. Although we considered and evaluated some initial criteria for ensuring general CA quality such as output formats, dialog control, or performance metrics (Lewandowski et al., 2023; Radziwill & Benton, 2017) in our initial design cycle, domain-independent quality attributes were not the focus.

User groups

In line with our stakeholder analysis, we noted different archetypes of CA users. Some participants heavily relied on the tool, while others used it as a backup. This aligns with the two proposed user levels (beginner and advanced) and matches observation on different interaction modes in CA usage in the workplace setting (e.g., Gkinko & Elbanna, 2023). Additionally, our findings indicate a significant difference in the variance of motivation between the treatment groups, indicating that the experience of the CA users differed strongly from user to user. This finding could emphasize the existence of different learning or GUI preferences. This was also reflected in the qualitative responses where participants expressing less satisfaction referred to three main reasons: either they found the CA to be somewhat circumstantial, preferred a more personal (i.e., human) support, or preferred other learning methods entirely.

Theoretical and practical contribution

With our research on designing a CA for data exploration in CS projects, we intend to advance the emerging practice of CS by guiding the designing of tools necessary for participation, thereby creating innovative and essential learning opportunities in data literacy. By going through and

reporting the different steps of the DSR process, we provide a series of contributions for theorists and practitioners.

Theoretical contributions

First, by connecting related literature on data literacy (e.g., D'Ignazio & Bhargava, 2016; Schüller et al., 2019), CS (e.g., Kloetzer et al., 2021; Tavanapour et al., 2019), and CAs (e.g., Okonkwo & Ade-Ibijola, 2021; Pérez et al., 2020), we provide valuable insights for future research projects at the intersection of these topics. Second, implementing a problem-centered design approach, we generate and evaluate design knowledge in the form of user needs and design principles on a CA assisting data exploration in the context of CS projects. Tailored to the inherent challenges of supporting CS projects and data exploration, the educational CA design presents a novel solution in this field. Thus, the contributed DSR knowledge can be classified as "Improvement" (Gregor & Hevner, 2013), and especially, DP1, DP2, and DP4 can be used as guidance for future design studies at the intersection of CAs, data literacy, and CS. In addition to descriptive principles, we also provide a design instantiation in the form of a prototypical CA (see Level 1, 2; Gregor & Hevner, 2013). In comparison to other research streams, giving insights in the design knowledge's usefulness is a central component in DSR (Peppers et al., 2020). In an experimental study, we demonstrated the positive effect of the CA on task performance in data exploration, which is a promising sign for CA technology in CS projects. The experiments' results, including the collected feedback and insights into user behavior, can be interesting starting points for further research.

Practical contributions

Based on our theoretic findings and our artifact instantiation, our work provides several interesting implications for CS practitioners and system developers. While our evaluated design principles can inform the development of new and individualized CS support tools for the context of data exploration, the open-source code of our artifact implementation can be directly used and adapted by interested practitioners. As such, we provide CS practitioners with a simple and low-effort opportunity to support citizen participation beyond data collection in the proceeding analysis phase, thereby realizing learning opportunities. Likewise, for policy-makers, we highlight how creative, non-formal teaching opportunities can be created and supported to address improving data literacy as a societal issue. Finally, beyond the CS use case, our CA design could be adopted by other stakeholders, such as educators or companies aiming to support their students or employees in improving their data exploration skills.

Limitations

As with any DSR study, the first cycle has limitations resulting from decisions made during the development of the artifact and the selected method for evaluation. Currently, the CA is limited in its outreach and applicability as it is only available in English and focuses exclusively on quantitative data, although it is intended to be dataset-independent. Regarding our approach to evaluation, we used an artificial use case instead of an actual CS project. This might have impacted participant motivation and perceptions of empowerment, as we envision such tools would enable actual citizen scientists to analyze data according to their inherent questions and ideas—rather than exogenous ones. Additionally, we observed participants' ages to be lower than they usually are for citizen scientists (Ciarán Mac Domhnaill & Nolan, 2020). Moreover, participants were extrinsically motivated to participate in the experiment through monetary compensation, which would not be true for citizen scientists. Therefore, while the study provides exciting insights into the usefulness of CAs for data exploration with citizens, it needs further consideration of the circumstances in CS projects and additional evaluation cycles to make more generalizable statements about their usefulness in this context.

Future work

This study's results and limitations give insights into possible future research strings. On the one hand, qualitative and quantitative feedback from the participants on the CA suggests a further refinement of its functionality and design. In this context, features that may enhance our artifact's hedonic quality should be tested. As a starting point, domain-independent quality attributes related to anthropomorphism, affect or accessibility (Lewandowski et al., 2023; Radziwill & Benton, 2017; Seeger et al., 2021) could be further considered for the specific use case of our artifact. On the other hand, further analyzing the lack of correlation between perceived learning and task performance and ways to circumvent this are interesting starting points for future research. For instance, it has been shown that feedback positively impacts students' perceived learning (e.g., Chan & Ko, 2021; Eom et al., 2006). Thus, it could be interesting to investigate whether the indication of the actual performance of a participant after the analysis task would change the results of the perceived learning and the correlation between the variables in our experiment. Additionally, it could be interesting to evaluate whether the Dunning-Kruger effect could explain the differences in perceived learning and task performance between the groups. Kruger and Dunning (1999) showed that people who lack knowledge or skills are likelier to be unaware of this in self-assessment. Instead, they overestimate their skills compared to participants with more

knowledge. Applied to our experiment, this could mean the amount of knowledge presented to the CA users could have negatively affected the user's assessment of aspects such as his or her feeling of empowerment or learning through the analysis activities, compared to a user not exposed to this knowledge. A third research string could focus on evaluating the CA in a real-life CS environment, including assessing who can use it and who is excluded. This perspective would be critical as CS strives for inclusivity (National Academies of Sciences & Medicine, 2018; Sorensen et al., 2019), and the under- or over-representation of particular groups can have negative consequences for project outcomes (Sorensen et al., 2019). Therefore, understanding who is excluded from our artifact and what alternatives could be created would be indispensable.

Conclusion

Inequalities in data access and literacy pose a risk to the individual and society as a whole. In this work, we therefore investigated the use case of CS as a means to empower citizens in accessing and working with data. We have presented results from the first cycle of a DSR project targeting the development of a CA to support data exploration in CS projects. Following the six steps for a problem-centered design process by Peffers et al. (2020), we have approached this challenge by structuring and elaborating on the problem space and its associated stakeholders, eliciting requirements and translating them into design principles for a solution artifact and finally presenting and testing a prototypical implementation of this artifact. The result of our first design cycle is a CA for data analysis activities offering flexible support on demand to multiple users in parallel. For inexperienced users, the tool provides seamless guidance through data analysis by providing dataset-independent knowledge and tips, allowing users to decide how deeply they want to dive into a particular topic. Advanced users can use a question-and-answer process to ask questions about data analysis freely, thus enriching only their knowledge with the CA and controlling the analysis process.

In its current state, the CA shows high potential for transferring required data literacy to citizens, enabling them to perform better in analytical tasks. Qualitative user feedback shows that multiple citizens perceive the tool's support as enjoyable and useful and point out potential application fields. Harnessing the advantages of their easy, resource-efficient provisioning, the usage of CAs in CS projects seems promising and could positively promote equitable access to data-driven knowledge. However, identified challenges, such as the participants' motivation, feeling of empowerment, and perceived learning effect, could not be solved adequately by

the CA. This indicates that further research is necessary to refine the CA and its usability, which we intend to accomplish in a proceeding design cycle.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12525-024-00705-3>.

Acknowledgements We want to thank the 12 workshop participants and the 30 experiment participants for their contribution to our work.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are available from the authors upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alaaeldin, R., Asfoura, E., Kassem, G., & Abdel-Haq, M. S. (2021). Developing chatbot system to support decision making based on big data analytics. *Journal of Management Information and Decision Sciences*, 24(2), 1–15.
- Alavi, M., Marakas, G. M., & Yoo, Y. (2002). A comparative study of distributed learning environments on learning outcomes. *Information Systems Research*, 13(4), 404–415. <https://doi.org/10.1287/isre.13.4.404.72>
- Anh, N., & Ngan, H. (2021). Artificial intelligence in mathematics education: An empirical study of using chatbot in teaching and learning mathematics at vietnamese high schools. *5th ASIA PACIFIC International Modern Sciences Congress*, 306–317.
- Athreya, R. G., NgongaNgomo, A.-C., & Usbeck, R. (2018). Enhancing community interactions with data-driven chatbots—the DBpedia chatbot. *The Web Conference 2018 - Companion of the World Wide Web Conference. WWW, 2018*, 143–146.
- Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education*, 70, 65–79. <https://doi.org/10.1016/j.compedu.2013.08.003>
- Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data* (Data-Pop alliance white paper series). <https://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data>. Accessed 23 Jan 2024.
- Bhargava, R., Kadouaki, R., Bhargava, E., Castro, G., & D'Ignazio, C. (2016). Data murals: Using the arts to build data literacy. *The*

- Journal of Community Informatics*, 12(3), 197–216. <https://doi.org/10.15353/joci.v12i3.3285>
- Bittner, E., Oeste-Reiß, S., & Leimeister, J. M. (2019). Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2019.035>
- Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Understanding of Science*, 25(1), 2–16. <https://doi.org/10.1177/0963662515607406>
- Bowser, A., Cooper, C., de Sherbinin, A., Wiggins, A., Brenton, P., Chuang, T.-R., Faustman, E., Haklay, M., & (Muki), & Meloche, M. (2020). Still in need of norms: The state of the data in citizen science. *Citizen Science: Theory and Practice*, 5(1), Article 1. <https://doi.org/10.5334/cstp.303>
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1481>
- Center for Self-Determination Theory. (2022). *Intrinsic Motivation Inventory (IMI)*. <https://selfdeterminationtheory.org/intrinsic-motivation-inventory>. Accessed 23 Jan 2024.
- Chan, S. C. H., & Ko, S. (2021). The dark side of personal response systems (PRSs): Boredom, feedback, perceived learning, learning satisfaction. *Journal of Education for Business*, 96(7), 435–444. <https://doi.org/10.1080/08832323.2020.1848769>
- Ciarán Mac Domhnaill, S. L., & Nolan, A. (2020). The citizens in citizen science: Demographic, socioeconomic, and health characteristics of biodiversity recorders in Ireland. *Citizen Science: Theory and Practice*, 5, 1–17. <https://doi.org/10.5334/cstp.283>
- Clarke, R. (2016) Big data big risks. Abstract. *Information Systems Journal*, 26(1), 77–90. <https://doi.org/10.1111/isj.12088>
- D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design principles, tools and activities for data literacy learners. *The Journal of Community Informatics*, 12(3), Article 3. <https://doi.org/10.15353/joci.v12i3.3280>
- D'Ignazio, C. (2022). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 6–18. <https://doi.org/10.1075/idj.23.1.03dig>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- De Albuquerque, J., & Almeida, A. (2020). Modes of engagement: Reframing 'sensing' and data generation in citizen science for empowering relationships. In T. Davies & A. Mah (Eds.), *Toxic truths: Environmental justice and citizen science in a post-truth age* (pp. 267–281). Manchester University Press.
- Debruyne, C., Kearns, A., O'Neill, C., Colclough, M., Grehan, L., & O'Sullivan, D. (2021). DALIDA: Data literacy discussion workshops for adults. *ACM International Conference Proceeding Series*, 23–25.
- Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, 23(1), 96–138. <https://doi.org/10.17705/1jais.00724>
- Ekinci, E., Omurca, S., & Acun, N. (2018). A comparative study on machine learning techniques using titanic dataset. In *7th International Conference on Advanced Technologies*, Antalya, Turkey (pp. 411–416).
- Eom, S. B., Wen, H. J., & Ashill, N. (2006). The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education*, 4(2), 215–235. <https://doi.org/10.1111/j.1540-4609.2006.00114.x>
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. The MIT Press.
- Fan, M., Li, Y., & Truong, K. N. (2020). Automatic detection of usability problem encounters in think-aloud sessions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 10(2), 1–24. <https://doi.org/10.1145/3385732>
- Gkinko, L., & Elbanna, A. (2023). The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users. *International Journal of Information Management*, 69, 102568. <https://doi.org/10.1016/j.ijinfomgt.2022.102568>
- Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, 48(10), 1159–1176. <https://doi.org/10.1002/tea.20442>
- Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, 43, 11–22. <https://doi.org/10.1111/test.12267>
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–356. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Gregor, S., Kruse, C. L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems*, 21(6), 1622–1652. <https://doi.org/10.17705/1jais.00649>
- Gupta, K., Sharma, P., & Bouza, C. N. (2018). Surviving the Titanic tragedy: A sociological study using machine learning models. *Suma De Negocios Magazine*, 9(20), 86–92. <https://doi.org/10.14349/sumneg/2018.v9.n20.a2>
- Hanley, H. W. A., & Durumeric, Z. (2023, May 16). Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites. ArXiv. Org. <https://arxiv.org/abs/2305.09820v1>
- Hinderks, A., Schrepp, M., & Thomaschewski, J. (2018). A benchmark for the short version of the user experience questionnaire. *Proceedings of the 14th International Conference on Web Information Systems and Technologies (WEBIST 2018)*, 373–377.
- Holowka, E. M., Woods, S., Pahayahay, A., Roy, M., & Khalili-Mahani, N. (2021). Principles for designing an mHealth app for participatory research and management of chronic pain. In V. G. Duffy (Ed.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management AI, Product and Service. Lecture Notes in Computer Science* (pp. 50–67). Springer, Cham.
- Hoon, G. K., Yong, L. J., & Yang, G. K. (2020). Interfacing chatbot with data retrieval and analytics queries for decision making. In A. Abdul Majeed, J. A. Mat-Jizat, M. H. A. Hassan, Z. Taha, H. L. Choi, & J. Kim (Eds.), *Rita 2018 - Proceedings of the 6th International Conference on Robot Intelligence Technology and Applications* (pp. 385–394). Springer Singapore.
- Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. Routledge.
- Isacco, S., Claps, P., Grasso, S., Ferrari, E., Guercio, M. B., Musumeci, R. E., Scarcella, G. E., Versace, P., & Laio, F. (2018). Floodbook: A social platform for flood hydrology. *Proceedings of the 13th International Conference on Hydroinformatics*, 3, 941–949.
- Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breiter, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. *Business and Information Systems Engineering*, 62(3), 211–225. <https://doi.org/10.1007/s12599-020-00644-1>
- Jayawickrama, T., Abdelaal, S., & Abadia, R. (2020). Using online learning environments to address digital literacy competencies of construction management graduates. *Proceedings of the 28th International Conference on Computers in Education, ICCE 2020*, 1, 670–679.

- Jennett, C., Kloetzer, L., Schneider, D., Iacovides, I., Cox, A., Gold, M., Fuchs, B., Eveleigh, A., Methieu, K., Ajani, Z., et al. (2016). Motivations, learning and creativity in online citizen science. *Journal of Science Communication*, 15(3). A05. <https://doi.org/10.22323/2.15030205>
- Jones, M. G., Childers, G., Andre, T., Corin, E. N., & Hite, R. (2018). Citizen scientists and non-citizen scientist hobbyists: Motivation, benefits, and influences. *International Journal of Science Education, Part B*, 8(4), 287–306. <https://doi.org/10.1080/21548455.2018.1475780>
- Keyner, S., Savenkov, V., & Vakulenko, S. (2019). Open data chatbot. In P. Hitzler, S. Kirrane, O. Hartig, V. de Boer, M.-E. Vidal, M. Maleshkova, S. Schlobach, K. Hammar, N. Lasierra, S. Stadtmüller, K. Hose, & R. Verborgh (Eds.), *The Semantic Web: ESWC 2019 Satellite Events* (pp. 111–115). Springer.
- Kim, H.-W., & Gupta, S. (2014). A user empowerment approach to information systems infusion. *IEEE Transactions on Engineering Management*, 61(4), 656–668. <https://doi.org/10.1109/TEM.2014.2354693>
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of American Society of Training Directors*, 13(3), 21–26.
- Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education*, 18(2), Article es2. <https://doi.org/10.1187/cbe.18-02-0023>
- Kloetzer, L., Lorke, J., Roche, J., Golubic, Y., Winter, S., & Jøgeva, A. (2021). Learning in citizen science. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 283–308). Springer.
- Krishnamurthy, R., & Awazu, Y. (2016). Liberating data for public value: The case of Data.gov. *International Journal of Information Management*, 36(4), 668–672. <https://doi.org/10.1016/j.ijinfomgt.2016.03.002>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kvale, K., Freddi, E., Hodnebrog, S., Sell, O. A., & Følstad, A. (2021). Understanding the user experience of customer service chatbots: What can we learn from customer satisfaction surveys? In A. Følstad, T. Araujo, S. Papadopoulos, E.L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design* (pp. 205–218). Springer.
- Lewandowski, T., Kučević, E., Leible, S., Poser, M., & Böhm, T. (2023). Enhancing conversational agents for successful operation: A multi-perspective evaluation approach for continuous improvement. *Electronic Markets*, 33, 39. <https://doi.org/10.1007/s12525-023-00662-3>
- Lia, E. H., Derrien, M. M., White, E. M., Winder, S. G., & Wood, S. A. (2023). Community science with a text-messaging chatbot provides a novel approach to monitoring outdoor recreation. *Digital Geography and Society*, 100059. <https://doi.org/10.1016/j.diggeo.2023.100059>
- Liu, H.-Y., Dörler, D., Heigl, F., & Grossberndt, S. (2021). Citizen science platforms. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 439–459). Springer.
- Logan, V. (2017). *Information as a second language: Enabling data literacy for digital society*. Gartner. <https://www.gartner.com/en/documents/3602517>.
- Matsuura, S., & Ishimura, R. (2017). Chatbot and dialogue demonstration with a humanoid robot in the lecture class. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Human and Technological Environments* (pp. 233–246). Springer, Cham.
- Monzón Alvarado, C. M., Zamora Rendon, A., & Vázquez Pérez, A. del S. (2020). Integrating public participation in knowledge generation processes: Evidence from citizen science initiatives in Mexico. *Environmental Science & Policy*, 114, 230–241. <https://doi.org/10.1016/j.envsci.2020.08.007>
- Narechania, A., Srinivasan, A., & Stasko, J. (2021). NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 369–379. <https://doi.org/10.1109/TVCG.2020.3030378>
- National Academies of Sciences, Engineering & Medicine. (2018). *Learning through citizen science: Enhancing opportunities by design*. The National Academies Press.
- Neumaier, S., Savenkov, V., & Vakulenko, S. (2017). Talking open data. *The Semantic Web: ESWC 2017 Satellite Events - ESWC 2017 Satellite Events*, 132–136.
- Nguyen, H. D., Pham, V. T., Tran, D. A., & Le, T. T. (2019). Intelligent tutoring chatbot for solving mathematical problems in high-school. *Proceedings of the 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–6.
- Okonko, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
- Paetsch, F., Eberlein, A., & Maurer, F. (2003). Requirements engineering and agile software development. *Proceedings of the 12th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, WET ICE 2003*, 308–313.
- Palan, S., & Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Panetta, K. (2021). *A data and analytics leader's guide to data literacy*. Gartner. <https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guideto-data-literacy>
- Peffer, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2020). Design science research process: A model for producing and presenting information systems research. *CoRR*, abs/2006.02763.
- Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549–1565. <https://doi.org/10.1002/cae.22326>
- Pettibone, L., Vohland, K., & Ziegler, D. (2017). Understanding the (inter)disciplinary and institutional diversity of citizen science: A survey of current practice in Germany and Austria. *PLoS ONE*, 12(6), 1–6. <https://doi.org/10.1371/journal.pone.0178778>
- Portela, M. (2021). Interfacing participation in citizen science projects with conversational agents. *Human Computation*, 8(2), 33–53. <https://doi.org/10.15346/hc.v8i2.114>
- Pullinger, J. (2021). Misuse of statistics: Time to speak out. *Statistical Journal of the IAOS*, 37(1), 79–84. <https://doi.org/10.3233/SJI-210783>
- Radermacher, W. J. (2021). Literacy in statistics for the public discourse. *Statistical Journal of the IAOS*, 37(3), 747–752. <https://doi.org/10.3233/SJI-210852>
- Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents (arXiv:1704.04579). [arXiv.https://doi.org/10.48550/arXiv.1704.04579](https://doi.org/10.48550/arXiv.1704.04579)
- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630. <https://doi.org/10.1016/j.ijhcs.2021.102630>
- Roberts, T. L., Lowry, P. B., & Sweeney, P. D. (2006). An evaluation of the impact of social presence through group size and the use of collaborative software on group member 'voice' in face-to-face and computer-mediated task groups. *IEEE Transactions on*

- Professional Communication*, 49(1), 28–43. <https://doi.org/10.1109/TPC.2006.870460>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461. <https://doi.org/10.1037/0022-3514.43.3.450>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103–108. <https://doi.org/10.9781/ijimai.2017.09.001>
- Schüller, K., Busch, P., & Hindinger, C. (2019). Future skills: A framework for data literacy. *Hochschulforum Digitalisierung*, 46, 1–128.
- Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with human-like conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 931–967. <https://doi.org/10.17705/1jais.00685>
- Shah, H. R., & Martinez, L. R. (2016). Current approaches in implementing citizen science in the classroom. *Journal of Microbiology & Biology Education*, 17(1), 17–22. <https://doi.org/10.1128/jmbe.v17i1.1032>
- Sharma, S., & Pandey, S. (2013). Revisiting requirements elicitation techniques. *International Journal of Computer Applications*, 75(12), 35–39.
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., & E, K. M., & Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. *Ecology and Society*, 17, 29–48. <https://doi.org/10.5751/ES-04705-17022>
- Simonofski, A., Zuiderwijk, A., Clarinval, A., & Hammedi, W. (2022). Tailoring open government data portals for lay citizens: A gamification theory approach. *International Journal of Information Management*, 65, 102511. <https://doi.org/10.1016/j.ijinfomgt.2022.102511>
- Simpson, R., Page, K. R., & De Roure, D. (2014). Zooniverse: Observing the world's largest citizen science platform. *Proceedings of the 23rd International Conference on World Wide Web*, 1049–1054. <https://doi.org/10.1145/2567948.2579215>
- Simud, T., Ruengittinun, S., Surasvadi, N., Sanglerdsinlapachai, N., & Plangprasopchok, A. (2020). A conversational agent for database query: A use case for thai people map and analytics platform. *Proceedings of the 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing, ISAI-NLP 2020*, pp 1–6. <https://doi.org/10.1109/ISAI-NLP51646.2020.9376833>
- Socientize. (2014). *Green paper on citizen science for Europe: Towards a society of empowered citizens and enhanced research*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research>
- Sorensen, A. E., Jordan, R. C., LaDeau, S. L., Biehler, D., Wilson, S., Pitas, J.-H., & Leisnham, P. T. (2019). Reflecting on efforts to design an inclusive citizen science project in West Baltimore. *Citizen Science: Theory and Practice*, 4(1).
- Spasiano, A., Grimaldi, S., Braccini, A. M., & Nardi, F. (2021). Towards a transdisciplinary theoretical framework of citizen science: Results from a meta-review analysis. *Sustainability*, 13(14), 7904.
- Spiers, H., Swanson, A., Fortson, L., Simmons, B., Trouille, L., Blickhan, S., & Lintott, C. (2019). Everyone counts? Design considerations in online citizen science. *Journal of Science Communication*, 18(1), A04. <https://doi.org/10.22323/2.18010204>
- Stieglitz, S., Hofeditz, L., Brünker, F., Ehnis, C., Mirbabaie, M., & Ross, B. (2022). Design principles for conversational agents to support emergency management agencies. *International Journal of Information Management*, 63, 102469. <https://doi.org/10.1016/j.ijinfomgt.2021.102469>
- Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The Ethnobot: Gathering ethnographies in the age of IoT. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, 1–13.
- Tamayo-Moreno, S., & Pérez-Marín, D. (2016). Adapting the design and the use methodology of a pedagogical conversational agent of secondary education to childhood education. *International Symposium on Computers in Education (SIIE), 2016*, 1–6. <https://doi.org/10.1109/SIIE.2016.7751819>
- 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*. Stockholm & Uppsala.
- Trzaskowski, J. (2022). Data-driven value extraction and human well-being under EU law. *Electronic Markets*, 32(2), 447–458. <https://doi.org/10.1007/s12525-022-00528-0>
- Twidale, M. B., Blake, C., & Gant, J. P. (2013). Towards a data literate citizenry. *ICConference 2013 Proceedings*. <https://doi.org/10.9776/13189>
- van der Goot, M. J., Hafkamp, L., & Dankfort, Z. (2021). Customer service chatbots: A qualitative interview study into the communication journey of customers. In A. Følstad, T. Araujo, S. Papadopoulos, E.L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *CONVERSATIONS 2020, 4th international workshop on chatbot research* (pp. 190–204). Springer.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- Vinales, G., Magnotta, S. R., Steffes, E., & Kulkarni, G. (2019). Description and evaluation of an innovative segmentation, targeting, and positioning activity using student perceived learning and actual student learning. *Marketing Education Review*, 29(1), 24–36. <https://doi.org/10.1080/10528008.2018.1493932>
- Wald, D. M., Longo, J., & Dobell, A. (2016). Design principles for engaging and retaining virtual citizen scientists. *Conservation Biology*, 30(3), 562–570. <https://doi.org/10.1111/cobi.12627>
- Watson, J., & Callingham, R. (2004). Statistical literacy: From idiosyncratic to critical thinking abstract. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education International Association for Statistical Education (IASE) Roundtable* (pp. 116–162). International Statistical Institute Voorburg.
- Wirtz, B. W., Weyerer, J. C., Becker, M., & Müller, W. M. (2022). Open government data: A systematic literature review of empirical research. *Electronic Markets*, 32(4), 2381–2404. <https://doi.org/10.1007/s12525-022-00582-8>
- Zhang, Y., Calyam, P., Joshi, T., Nair, S., & Xu, D. (2019). Domain-specific topic model for knowledge discovery through conversational agents in data intensive scientific communities. *Proceedings of 2018 IEEE International Conference on Big Data, Big Data 2018*, 4886–4895

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.