Journal of the Association for Information Systems

Volume 24
Issue 6 *Special Issue: Data Science for Social Good (pp. 1439-1458, 1479-1499) Special Issue: Envisioning Digital Transformation: Advancing Theoretical Diversity (pp. 1459-1478, 1594-1798)*

Article 7

2023

Designing Conversational Dashboards for Effective Use in Crisis Response

Marcel Ruoff
Karlsruhe Institute of Technology, marcel.ruoff@kit.edu

Ulrich Gnewuch
Karlsruhe Institute of Technology, ulrich.gnewuch@kit.edu

Alexander Maedche
Karlsruhe Institute of Technology, alexander.maedche@kit.edu

Benjamin Scheibehenne
Karlsruhe Institute of Technology, scheibehenne@kit.edu

Recommended Citation

Ruoff, Marcel; Gnewuch, Ulrich; Maedche, Alexander; and Scheibehenne, Benjamin (2023) "Designing Conversational Dashboards for Effective Use in Crisis Response," *Journal of the Association for Information Systems*, 24(6), 1500-1526.
DOI: 10.17705/1jais.00801
Available at: https://aisel.aisnet.org/jais/vol24/iss6/7

Journal of the Association for Information Systems (2023) **24**(6), 1500-1526
doi: 10.17705/1jais.00801

RESEARCH ARTICLE

ISSN 1536-9323

Designing Conversational Dashboards for Effective Use in Crisis Response

Marcel Ruoff,[1] Ulrich Gnewuch,[2] Alexander Maedche,[3] Benjamin Scheibehenne[4]
[1]Karlsruhe Institute of Technology, Germany, marcel.ruoff@kit.edu
[2]Karlsruhe Institute of Technology, Germany, ulrich.gnewuch@kit.edu
[3]Karlsruhe Institute of Technology, Germany, alexander.maedche@kit.edu
[4]Karlsruhe Institute of Technology, Germany, scheibehenne@kit.edu

Abstract

Governments and health organizations are increasingly using dashboards to provide real-time information during natural disasters and pandemics. Although these dashboards aim to make crisis-related information accessible to the general public, the average user can have a hard time interacting with them and finding the information needed to make everyday decisions. To address this challenge, we draw on the theory of effective use to propose a theory-driven design for conversational dashboards intended for crisis response that can improve users' transparent interaction with these dashboards and facilitate access to crisis-related information during crises. We instantiate our proposed design in a conversational dashboard for the COVID-19 pandemic that enables natural language interaction in spoken or written form and helps users familiarize themselves with the use of natural language through conversational onboarding. The evaluation of our artifact shows that being able to use natural language improves users' interaction with the dashboard and ultimately increases their efficiency and effectiveness in finding information. This positive effect is amplified when users complete the onboarding before interacting with the dashboard, particularly when they can use both natural language and mouse interaction. Our findings contribute to research on dashboard design, both in general and in the specific context of crisis response, by providing prescriptive knowledge for extending crisis response dashboards with natural language interaction capabilities. In addition, our work contributes to the democratization of data science by proposing design guidelines for making information on crisis response dashboards more accessible to the general public.

Keywords: Dashboard, Conversational User Interface, Crisis Response, Design Science Research, Theory of Effective Use

Roger Chiang, Ahmed Abbasi, and Jennifer Xu were the accepting senior editors. This research article was submitted on April 15, 2021 and underwent two revisions. This paper is part of the Special Issue on Data Science for Social Good.

1 Introduction

By nature, crises are unpredictable, sudden, and often chaotic situations. When a crisis occurs, people want to find accurate and up-to-date information quickly so that they can make the best decisions for themselves, their families, and their communities (Leong et al., 2015). To satisfy information needs, governments and health organizations are increasingly relying on crisis response dashboards. Similar to business intelligence (BI) dashboards designed to support decision makers in organizations (Abbasi et al., 2016), crisis response dashboards are designed to provide citizens with key information about the current state of a crisis. As such, these data science artifacts primarily aim to democratize data science by making complex data accessible to the general public (Koch, 2021; Matheus et al., 2020). While crisis response dashboards have previously been developed for earthquakes (Zook et al., 2010), wildfires (Liu & Palen, 2010), and virus outbreaks (Cheng et al., 2011), they took center stage during the COVID-19 pandemic (Pietz et al., 2020). For example, the dashboard provided by Johns Hopkins University received more than a billion hits per day during the height of the pandemic (Gardner et al., 2021). COVID-19 dashboards served not only as a primary source of information about cases, deaths, and other key metrics for the general public, but were also frequently used to guide everyday decision-making (e.g., about visiting a friend or getting a haircut) (Flowers, 2020).

Both researchers and practitioners share the assumption that COVID-19 dashboards were highly effective in helping billions of users find the information they needed quickly, as long as the underlying data was accurate and the visualizations were interactive (Patino, 2021; Soper et al., 2021). However, while the reported numbers of daily dashboard users certainly look impressive, we know from the literature that to achieve their goals, people must use information systems (IS) effectively—rather than just using them (Burton-Jones & Grange, 2013). The fundamental dimension of effective use is *transparent interaction*, which describes how well users can access information from an IS unimpeded by its physical and surface structures (e.g., the user interface) (Burton-Jones & Grange, 2013). If users are unable to interact with a dashboard transparently, they are unlikely to find the information they need and make good decisions (e.g., about wearing a mask in regions with increasing case numbers).

Against this backdrop, it is important to highlight that achieving transparent interaction with dashboards in general, and crisis response dashboards in particular, may be more difficult than expected, especially for users who are not familiar with the technology and/or have limited domain knowledge. Such users often struggle to obtain the needed information due to the complexity of the dashboard and difficulties navigating its interface (Young et al., 2021; Young & Kitchin, 2020). Additionally, anecdotal evidence from a review of 52 state-level COVID-19 dashboards in the United States shows that many of them “were overly complex to navigate, and even experienced health researchers had difficulty finding key information” (Prevent Epidemics, 2020, p. 17). These findings suggest that users often face difficulties in transparently interacting with such dashboards and that finding the information they need quickly may thus not be as easy as intended.

Given that crisis response dashboards, such as the ones developed for the COVID-19 pandemic, are designed to inform the general public, it is imperative that they enable a wide range of users—regardless of their sociodemographic backgrounds and technical expertise—to achieve transparent interaction. A promising way to address this challenge is to move beyond the traditional graphical user interface (GUI) and provide users with a more natural way of interacting with such dashboards using natural language. With recent technological advances in artificial intelligence (AI), natural language could make navigating such dashboards and finding information less difficult because it allows users to articulate their information needs more naturally, as they would in everyday conversation (Lee et al., 2020). However, despite these technological advances, we know little about how to design a crisis response dashboard with natural language interaction capabilities (hereafter referred to as a *conversational dashboard*) and do not yet understand whether natural language would actually enable users to interact with these dashboards more transparently. Therefore, we seek to answer the following research question:

RQ: How can crisis response dashboards be extended with natural language interaction capabilities to improve users’ transparent interaction and access to crisis-related information?

To address this question, we follow the design science research (DSR) approach (Hevner et al., 2004). Drawing on Burton-Jones & Grange’s (2013) theory of effective use (TEU), we propose a theory-driven design for conversational dashboards intended for crisis response and instantiate our proposed design in a novel data science artifact: a conversational dashboard for the COVID-19 pandemic that enables natural language interaction in spoken or written form and helps users familiarize themselves with the use of natural language through conversational onboarding. The evaluation of our artifact shows that the ability to use natural language improves users’ transparent interaction with the dashboard and ultimately increases their efficiency and effectiveness in finding the information they need. Our artifact thus contributes to the democratization of data science in the context of crisis response by making the information on dashboards more accessible to broader audiences, thereby narrowing the gap between data and insights. Our work also contributes to research on dashboard design and use, both in general and in the specific context of crisis response, by providing prescriptive knowledge for extending dashboards with natural language interaction capabilities. In addition, our findings shed light on potential design trade-offs that surface when users are provided with multiple ways of interacting with a dashboard and suggest an approach for addressing these trade-offs using conversational onboarding. Our findings provide actionable guidance to data scientists and dashboard providers on how to design crisis response dashboards that are more accessible to broader audiences.

2 Theoretical Foundations and Related Work

Our work is situated at the intersection of two research streams: conversational user interfaces (CUIs) and dashboards. Here, we first provide an overview of related work in these streams from both an IS and a human-computer interaction (HCI) perspective. Next, we describe existing research at the intersection of CUIs and dashboards (i.e., on conversational dashboards), which has emerged as a prominent research area in the HCI field. Finally, we introduce our kernel theory (i.e., TEU) and explain its key constructs.

2.1 Conversational User Interfaces

Conversational user interfaces (CUIs) enable people to interact with IS using spoken or written language in a natural way. The term *conversational* specifically emphasizes that these interfaces support the use of spontaneous natural language, in contrast to earlier applications (e.g., interactive voice response systems) that required a more restricted form of user input (e.g., “Press or Say 1 for English”) (McTear et al., 2016). In recent years, CUIs in the form of chatbots and conversational agents have attracted considerable interest from IS researchers (Diederich et al., 2022). A key focus of this research is the empirical investigation of how the human-like design of CUIs influences user perceptions and behaviors (e.g., Gnewuch et al., 2022; Schanke et al., 2021; Seeger et al., 2021). Further, prior IS studies have focused on designing CUIs for specific contexts, such as border screening (Nunamaker et al., 2011), job interviews (Diederich et al., 2020), or mental health care (Ahmad et al., 2022).

Additionally, the HCI field has a long tradition of investigating CUI design, dating back to the 1960s when the first chatbot, ELIZA, was developed (Weizenbaum, 1966). A key focus in this research stream is the examination of users’ expectations of and interactions with CUIs in real-life settings in order to identify design challenges (e.g., Luger & Sellen, 2016; Porcheron et al., 2018). For example, Luger and Sellen (2016) found that users often do not understand the limitations of CUIs and therefore need to be given feedback about the actual capabilities. A related line of research seeks to address the challenges related to ambiguity and complexity in natural language interaction. For example, existing studies have suggested design principles for handling conversational breakdowns (Ashktorab et al., 2019; Ruoff et al., 2022) and for providing conversational context to help users interact with CUIs (Jain et al., 2018). Another, more technical set of studies in this stream focuses on the development of new system architectures and the application of advanced machine learning techniques to improve the technical components underlying a CUI (e.g., Huang et al., 2018; Xu et al., 2017). Finally, a growing number of studies have investigated the design of CUIs for specific contexts (e.g., virtual team collaboration; Benke et al., 2020) and specific target groups (e.g., children; Zhang et al., 2022).

A great deal of research in both IS and HCI has addressed CUIs as an alternative to graphical user interfaces (GUIs). Several tech companies have even claimed that it is only a matter of time before CUIs replace apps and websites equipped with GUIs (McTear et al., 2016). However, it is difficult to convey the amount of visual information rich GUIs provide—e.g., data visualization in a dashboard—using natural language. This suggests that more could be achieved by complementing rather than replacing a GUI with a CUI. Against this backdrop, we next introduce related work on dashboards that typically feature rich GUIs and then present prior research on conversational dashboards that aim to combine both types of user interface.

2.2 Dashboards

Dashboards are “visual displays of the most important information needed to achieve one or more objectives; consolidated and organized on a single screen so the information can be monitored at a glance” (Few, 2006). Many organizations use BI dashboards to provide decision makers with a comprehensive overview of key performance indicators, thereby supporting their decision-making (Abbasi et al., 2016; Chen et al., 2012). Against this backdrop, most IS studies focus on dashboards designed for domain experts in organizations. Examples include business users in areas such as supply chain management (Park et al., 2016) and health professionals such as physicians (Chen et al., 2016). While the specific contexts and dashboard designs may differ, the target users in these studies are all familiar with the application domain, which helps them understand the data underlying the dashboard, and are likely to use the dashboard on a regular basis as part of their job. In contrast, very little IS research has been devoted to dashboards designed for broader audiences outside of organizational structures, who might be less familiar with dashboard technology. Thus, existing dashboard designs rarely include additional integrated learning features beyond help buttons (e.g., Nadj et al., 2020; Nguyen et al., 2021) or tooltips (e.g., Vallurupalli & Bose, 2018), which would benefit such audiences in particular. Recker’s (2021) study is the only one that focuses on the general public as target users of a dashboard and it is also one of the few IS studies that investigate dashboards in the context of crisis response. Overall, this dearth of research is surprising given the increasing pervasiveness of dashboards designed for broader audiences, particularly in the crisis response context (Koch, 2021; Matheus et al., 2020).

Further, existing dashboards found in the IS literature almost exclusively rely on GUIs to display data visualizations, ranging from simple line charts (e.g., Nguyen et al., 2021) to more complex network graphs (e.g., Lu et al., 2021). These dashboards typically provide additional features, such as filters and drill-downs, to enable users to interact with visualizations and navigate the GUI. While GUIs are well suited to display complex data visualizations, research suggests that users who are not familiar with dashboards and have limited domain knowledge may struggle to interact with them (Young et al., 2021). Therefore, other types of user interfaces (e.g., CUIs) might be more suitable for less tech-savvy audiences (Lee et al., 2020). Thus far, however, no IS study has investigated a dashboard with a CUI.

2.3 Conversational Dashboards

In contrast to the IS literature focused on investigating dashboards equipped with traditional GUIs, HCI research has considered CUIs as a promising extension to make dashboards more accessible (Lee et al., 2020). A key focus of this research is providing and improving the technical foundations that enable natural language interaction with data visualizations on a conversational dashboard. For example, several studies address the challenges of ambiguity in natural language by proposing design features for disambiguating unclear user input (e.g., Gao et al., 2015; Setlur et al., 2016). In addition, an emerging body of work explores how users interact with conversational dashboards using speech, touch, and keyboard input (e.g., Ruoff & Gnewuch, 2021; Saktheeswaran et al., 2020).

However, similar to IS research, most HCI studies have focused on dashboards designed for domain experts and tech-savvy groups of users (e.g., data analysts or computer science students; Gao et al., 2015; Setlur et al., 2016). The only study that specifically targeted the general public is a study that developed a smartphone app for exploring personal health data captured by a Fitbit tracker (Kim et al., 2021). Further, prior HCI research has predominantly focused on assessing the practical viability of conversational dashboards using relatively small samples (Srinivasan et al., 2020) rather than conducting rigorous evaluations of the underlying design principles. For example, Setlur et al. (2016) compared their conversational dashboard to a traditional dashboard without CUI in a user study with 12 domain experts from a software company.

Based on our review of the IS and HCI literature, we make three major observations about the current state of research on dashboard design. First, as the literature has primarily focused on dashboards designed for domain experts within organizational settings (e.g., managers, physicians) or for tech-savvy user groups (e.g., data analysts), research on the design of crisis response dashboards for broader audiences is scarce. This gap in the literature requires attention because previous studies have indicated that novice and less tech-savvy users particularly can find interacting with a dashboard difficult (Young et al., 2021; Young & Kitchin, 2020), suggesting that a different dashboard design is needed to accommodate broader audiences. Second, although HCI research has identified CUIs as a promising way to make dashboards more accessible, existing designs of conversational dashboards have not been derived from a solid theoretical foundation and are often not rigorously evaluated to ensure their utility. As a result, a theory-driven design for conversational dashboards, particularly for the crisis response context, is lacking. Finally, while research on the development of advanced dashboard features (e.g., new visualizations, better analytical capabilities) has prospered, much less research has examined integrated learning features that would particularly benefit the average user who is less familiar with dashboards and how to use natural language to interact with them. This is another critical research gap since users of crisis response dashboards have likely not received any dedicated training and do not have an IT department to consult for assistance. Our work addresses these gaps in the literature by proposing, instantiating, and rigorously evaluating a theory-driven design for conversational dashboards intended for crisis response that improves users' interaction with such dashboards and facilitates access to crisis-related information.

2.4 Theory of Effective Use

Drawing on representation theory, Burton-Jones and Grange (2013) proposed TEU based on the premise that rather than just being used, IS must be used effectively to obtain maximum benefits from them. They defined effective use as "using a system in a way that helps attain the goals for using the system" (p. 633) and conceptualized it as an aggregate construct with three hierarchical dimensions: (1) transparent interaction, (2) representational fidelity, and (3) informed action. This paper focuses on the first dimension of effective use—transparent interaction. According to Burton-Jones and Grange (2013), each lower-level dimension is necessary but not sufficient for the higher-level dimension. Therefore, when users are unable to interact with an IS transparently (transparent interaction), their chances of obtaining faithful representations (representational fidelity) and eventually acting upon these representations in an informed way (informed action) are dramatically reduced, if not eliminated. Transparent interaction is formally defined as "the extent to which a user is accessing the system's representations unimpeded by its surface [e.g., user interface] and physical structure [e.g., computer, input/output devices]" (p. 642). For example, the surface structure of a traditional dashboard is a GUI, which typically consists of menus, sliders, and additional interactive features that can be used to navigate the GUI and change data visualizations.

TEU also identifies two major factors that act as drivers of effective use: adaptation and learning (Burton-Jones & Grange, 2013). Adaptations are users' actions to improve the representations in a system or the way they can be accessed (e.g., through the surface structure). Learning involves users' actions to learn the system's components (e.g., representations, surface structure), the fidelity of its representations, and how to leverage representations toward taking more informed action. Given our emphasis on transparent interaction, we focus on two specific adaptation and learning actions that can increase users' ability to interact with a system transparently, namely adapting surface structure and learning surface structure. Typically, users can engage in adapting a system's surface structure by personalizing the user interface themselves or by suggesting improvements to system designers, who then adapt the interface for them (Barki et al., 2007). In addition, organizations introducing new IS usually offer training sessions and provide system manuals to facilitate users' learning of the system's surface structure (Lauterbach et al., 2020). However, in the context of crisis response dashboards, such strategies would be difficult to implement because these dashboards are often used in an ad hoc manner and, unlike in an organization, there is no clearly defined group of users. Consequently, TEU as a kernel theory provides convincing theoretical arguments on *why* adaptation and learning should improve transparent interaction, but it does not offer prescriptive guidance on *what* should be done through design to address users' lack of transparent interaction nor on *how* to achieve this. Therefore, design knowledge on how to adapt the surface structure of a crisis response dashboard and facilitate users' learning to improve transparent interaction is scarce.

3 Designing Conversational Dashboards for Crisis Response

Our research project follows the DSR approach (Hevner et al., 2004) to design a conversational dashboard for crisis response that improves users' transparent interaction and access to crisis-related information. The DSR approach is well-suited to guide our research, as it aims to generate design knowledge through innovative solutions for real-world problems (Hevner et al., 2004). In this section, we first describe our design process and then elaborate on the design outcomes, that is, our meta-requirements (MRs), design principles (DPs), and the software artifact.

3.1 Design Process

We adopted the DSR framework proposed by Kuechler and Vaishnavi (2008) and divided our project into two iterative build-evaluation cycles. Here, we briefly summarize our activities in each cycle. As illustrated in Figure 1, the work presented in this paper primarily focuses on the outcomes of the second and final design cycle.

We started the *first cycle* by gaining an in-depth understanding of the problem space in order to identify barriers and design challenges that make it difficult for broader audiences to interact with crisis response dashboards. In this step, we first conducted a review of the IS and HCI literature on the design and use of dashboards in several application areas including, but not limited to, crisis response. To supplement what we found in the literature, we conducted interviews with six actual and potential dashboard users (three women, three men) with an average age of 53.2 years (SD = 23.2) and diverse backgrounds (e.g., seniors, students, professionals).

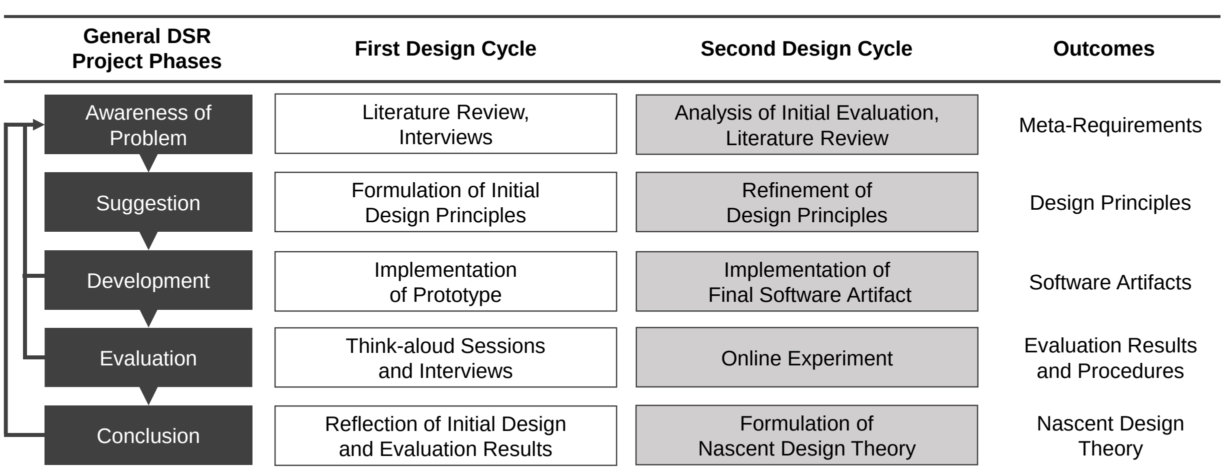

Figure 1. Overview of our DSR Approach

Our goal was not to obtain as representative a sample as possible but rather to include less tech-savvy participants who do not use dashboards on a regular basis. In the interviews, we encouraged participants to interact with the Johns Hopkins University COVID-19 dashboard (Dong et al., 2020) and then asked them about the challenges they faced during the interaction. The findings from the interviews and the literature review revealed that transparent interaction with the dashboard is particularly important for effective use but achieving it can be more difficult than expected. Drawing on TEU (Burton-Jones & Grange, 2013) as our overarching kernel theory, we then derived two MRs. Subsequently, we proposed three initial DPs for conversational dashboards to address these MRs based on the idea that natural language interaction can help users achieve higher levels of transparent interaction with a dashboard. Finally, we instantiated our initial DPs in a first prototype that had natural language interaction capabilities but, in contrast to our final artifact, did not yet offer conversational onboarding. Instead, we implemented both a help button and a help message in the chat to provide instructions on how to interact with the dashboard. We evaluated our prototype with fifteen participants (seven women, eight men) with an average age of 43.1 years (*SD* = 22.7) and different levels of IT experience using think-aloud sessions combined with interviews. Overall, we found that all participants appreciated being able to use natural language to interact with the dashboard. Less tech-savvy participants reported that it allowed them to directly formulate their information needs in natural language and navigate the dashboard without dealing with interactive features such as sliders or filters. Conversely, more tech-savvy participants highlighted that using natural language improved their efficiency in the interaction and allowed faster access to the information on the dashboard. Nonetheless, most participants stated that they would prefer natural language as an addition to rather than a replacement for mouse interaction. Additionally, we found that most participants did not use or even recognize the help button or help message. Consequently, one of the key challenges participants mentioned was their lack of familiarity with and confidence in using natural language to interact with the conversational dashboard. This finding showed that our initial design was unable to provide sufficient support for users in learning how to interact with the dashboard, thus highlighting the need to refine our DPs to better address the MRs in the final artifact. Consequently, this reflection served as the entry point to the second cycle and eventually led to the development of the conversational onboarding.

The *second cycle* started with a refinement of the initial problem definition, MRs, and DPs. Based on the results of the first evaluation, we realized that users need a more systematic, "hands-on" approach to learn how to use natural language to interact with the dashboard. Therefore, we extended our review of dashboard studies and specifically analyzed the design of integrated learning features in current dashboards. Since the results showed that most dashboards rely on help buttons and tooltips (similar to our first prototype), we took inspiration from research on technology-mediated learning that has proposed the concept of enactive learning for enhanced learning outcomes (Gupta et al., 2010; Gupta & Bostrom, 2009). Drawing on this concept, we then refined our third DP based on the idea of conversational onboarding. Subsequently, we developed a fully functional version of our artifact that instantiated the refined DPs. To rigorously evaluate the DPs, we conducted a large-scale online experiment with 271 participants and measured their level of transparent interaction. Finally, we abstracted and synthesized the design and evaluation results into a nascent design theory for conversational dashboards intended for crisis response.

3.2 Problem Description and Meta-Requirements

Crisis response dashboards, such as the ones developed for the COVID-19 pandemic, are designed to give the general public access to important information during a crisis (Ivanković et al., 2021; Recker, 2021). However, our review of the dashboard literature in IS and HCI and our interviews with less tech-savvy individuals suggest that the average user is likely to have a hard time interacting with a crisis response dashboard and ultimately finding the needed information. For example, one interviewee mentioned that she had to "search the dashboard extensively before even knowing how to get to the needed information." Another interviewee explained that he "did not know what changed in the visualization based on [his] interaction." Viewed through the lens of our kernel theory, there often appears to be a lack of transparent interaction with crisis response dashboards.

To derive meta-requirements (MRs) for addressing this problem, we draw on TEU (Burton-Jones & Grange, 2013). As outlined in Section 2.4, TEU proposes two important factors that can improve transparent interaction: adaptation and learning. Given that the design problem we address relates to the difficulties in interacting with the user interface of crisis response dashboards, we specifically focused on TEU's adaptation and learning actions related to surface structure. Drawing on these theoretical underpinnings, we derive two MRs on how the dashboards' surface structure might be adapted and how learning it could be better supported.

In line with TEU, *adapting* the dashboard's surface structure (i.e., its user interface) is one approach to improving users' transparent interaction with the dashboard and ultimately their access to information.

The surface structure of current crisis response dashboards consists of a GUI that can primarily be navigated using a mouse, keyboard, or touchscreen. Therefore, a promising way to adapt the surface structure is to move beyond the traditional GUI and provide users with a more natural way of interacting with the dashboard, for example, using natural language (Lee et al., 2020). Natural language could simplify the dashboard navigation and thus make finding information less difficult by allowing users to formulate their information needs more naturally, as they would in everyday conversation. One interviewee hinted at this possibility by wondering "why [he] could not just ask the dashboard and talk to it." Following this line of thought, we propose our first MR:

MR1: The surface structure of a crisis response dashboard should be adapted to allow for a more natural way of interaction in order to improve transparent interaction.

A second, complementary approach to improve users' transparent interaction with a crisis response dashboard would be to support users in *learning* how to interact with the surface structure of a crisis response dashboard (Burton-Jones & Grange, 2013). In contrast to dashboard users in organizations (e.g., managers, health professionals), the average user of a crisis response dashboard would likely not have received any dedicated training and not be able to call an IT department for assistance. Since current crisis response dashboards primarily offer integrated learning features in the form of passive help buttons and tooltips, a promising way to facilitate users' learning of its surface structure would be to enable the dashboard to actively familiarize users with possible ways of interaction, particularly when it offers novel ways of interacting with which users may not be familiar (e.g., using natural language). Based on these considerations, we propose our second MR:

MR2: A crisis response dashboard should actively support users in independently learning its surface structure in order to improve transparent interaction.

3.3 Design Principles

To address the two identified MRs, we derived three DPs by building on existing theory and the current body of prescriptive knowledge for dashboards. Regarding our first MR, namely adapting the dashboard's surface structure to enable a more natural way of interaction, we drew on the concept of affordances (Gibson, 1977), which is linked to TEU in several ways (Burton-Jones & Volkoff, 2017). Affordances are a key concept in the HCI and IS fields to describe and understand how users interact with an IS, thereby providing a solid theoretical grounding for our first and second DPs.

Affordances are defined as action possibilities that the environment provides to an actor (Gibson, 1977). According to Burton-Jones and Grange (2013), the surface structure of an IS relates particularly to physical affordances. Physical affordances are design features, such as buttons, that help users perform a physical action on the user interface (Hartson, 2003). For example, dashboards offer interactive features, such as menus, sliders, and filters, that enable users to directly change the data visualizations. However, actualizing these physical affordances is difficult for some users, for example, because they do not know how and when to use the interactive features that enable navigating the dashboard. To address this challenge and offer users a more natural way of finding the information they need, we propose using natural language, which is the primary means of communication between humans (Knote et al., 2021). In contrast to clicking buttons, scrolling, and setting filters, natural language can provide a more natural way of performing actions on the interface and therefore "make affordances easy to actualize" (Knote et al., 2021, p. 434). It might also require less effort because users could directly use natural language input instead of translating their information needs into a series of actions on the interface (e.g., setting filters). While the possibility of having a natural conversation with a dashboard might have seemed far-fetched in the past, recent technological advances, particularly in the area of large generative language models (e.g., GPT-3), suggest that in the future developers can make this scenario a reality with minimal manual effort or domain knowledge. Consequently, we propose enabling users to seamlessly navigate the dashboard using natural language. Thus, based on MR1, we formulate our first DP using the schema suggested by Gregor et al. (2020):

DP1: To enable the general public to seamlessly navigate a dashboard for crisis response, provide users with the ability to use spoken or written language in a natural way because articulating an information need in natural language is easier than translating it into a series of actions on the graphical user interface.

While the first DP postulates that a crisis response dashboard should allow natural language interaction, it does not specify whether the ability to use natural language should complement or replace existing ways of interacting with a dashboard (e.g., using a mouse). At first glance, it could seem better to restrict users to natural language interaction, thereby removing the need for them to understand how and when to use interactive features, such as menus, sliders, and filters, to navigate the dashboard. However, according to TEU (Burton-Jones & Grange, 2013), transparent interaction involves not only the system itself (e.g., a dashboard) but also the user and task. This clarification is particularly important for crisis response dashboards because they need to accommodate a wide range of users, ranging from novices who have never seen a dashboard to tech-savvy groups of individuals (Ivanković et al., 2021). Therefore, if different users prefer different ways of interaction to achieve the same goal, restricting them to natural language only could backfire. Additionally, the characteristics of the task at hand, such as its complexity, can also influence the suitability of using natural language or a mouse for a particular task. Considering this, we argue that users should be able to use both natural language and a mouse and need the freedom to choose between them in their interaction with the dashboard. Thus, based on MR1, we formulate our second DP:

DP2: To enable the general public to seamlessly navigate a dashboard for crisis response, provide users with the ability to choose between natural language and mouse interaction because it gives them flexibility for the task at hand and takes their individual preferences into account.

Our second MR focuses on supporting users in independently learning the surface structure of a crisis response dashboard. To formulate our third DP based on MR2, we drew on the concept of enactive learning (Gupta & Bostrom, 2009). Enactive learning has proven to be a feasible approach for web-based training and therefore provides a good theoretical foundation for addressing MR2, particularly because formal training approaches are difficult if not impossible to implement in the context of crisis response dashboards designed for the general public. As enactive learning "involves learning from the consequences of one's actions" (Gupta et al., 2010, p. 16), it is an effective approach to onboard users to a new IS. Based on the idea of providing "a guided simulated environment with rich feedback to [enable users to] evaluate their actions" (Gupta et al., 2010, p. 18), we propose integrating conversational onboarding that allows users to familiarize themselves with using natural language to interact with the crisis response dashboard. Given the relative novelty of natural language interaction, particularly in the context of crisis response dashboards, conversational onboarding should provide users with the opportunity to try out interacting with the dashboard using natural language in a step-by-step manner, observe the consequences of their actions (e.g., how and why data visualizations change), and receive feedback when something goes wrong. Then, before actually using the dashboard to find the information they are looking for, users can learn how to use natural language to navigate the dashboard. Taken together, based on MR2, we formulate our third DP as follows:

DP3: To enable the general public to seamlessly navigate a conversational dashboard for crisis response, provide users with conversational onboarding that takes them step-by-step through the natural language interaction with the dashboard because this helps users familiarize themselves with how to interact with the dashboard using spoken or written language.

3.4 Testable Design Propositions

Testable design propositions are a core component of a design theory (Gregor & Jones, 2007). Through the lens of our kernel theory, we therefore derived two testable propositions from the presented DPs. The primary outcome of interest and core construct from TEU is transparent interaction, which can be understood as the extent to which users can access information from an IS unimpeded by its user interface (Burton-Jones & Grange, 2013). As noted earlier, users can struggle to navigate the rich GUI of current crisis response dashboards to the information they need, due to difficulties with dashboards' sliders, filters, and other interactive features. Considering these challenges, we argue that users can interact with a crisis response dashboard more transparently if they have the ability to choose spoken or written language in their navigation of the dashboard. Compared with users having to translate an information need into a series of actions on the GUI (e.g., button clicks), which requires knowing and being able to use its features, formulating it in natural language would be much easier (Lee et al., 2020). Consequently, users should be able to achieve higher levels of transparent interaction with a conversational dashboard built according to our DP1 and DP2. Hence, we propose:

Proposition 1: A crisis response dashboard equipped with a conversational user interface allowing users to interact with the dashboard using natural language will enable them to achieve higher levels of transparent interaction.

TEU also posits that learning how to interact with the user interface of an IS can improve transparent interaction. As described earlier, current crisis response dashboards offer little help beyond a few tooltips and help buttons in teaching users how to navigate the interface and access information. Moreover, natural language is a rather new form of dashboard interaction that users might still need to learn. Therefore, we argue that providing users with conversational onboarding that can walk them through natural language interaction with the dashboard (DP3) should facilitate their learning by helping users familiarize themselves with using spoken or written language to navigate the dashboard. Consequently, users should be able to achieve higher levels of transparent interaction if the conversational dashboard offers conversational onboarding built according to DP3. Hence, we propose:

Proposition 2: A conversational crisis response dashboard equipped with conversational onboarding that walks users through natural language interaction with the dashboard will enable them to achieve higher levels of transparent interaction.

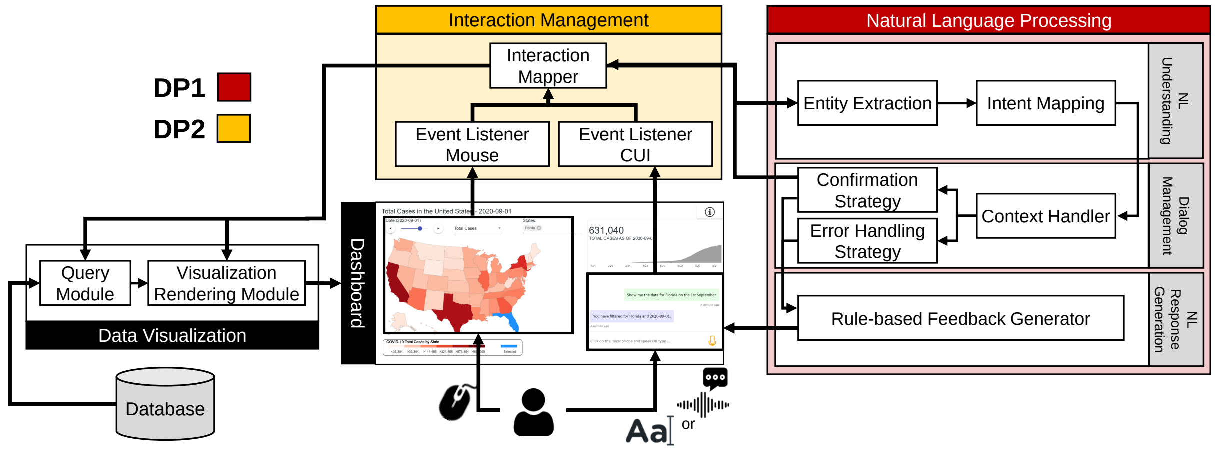

Figure 2. System Architecture of the Conversational Dashboard

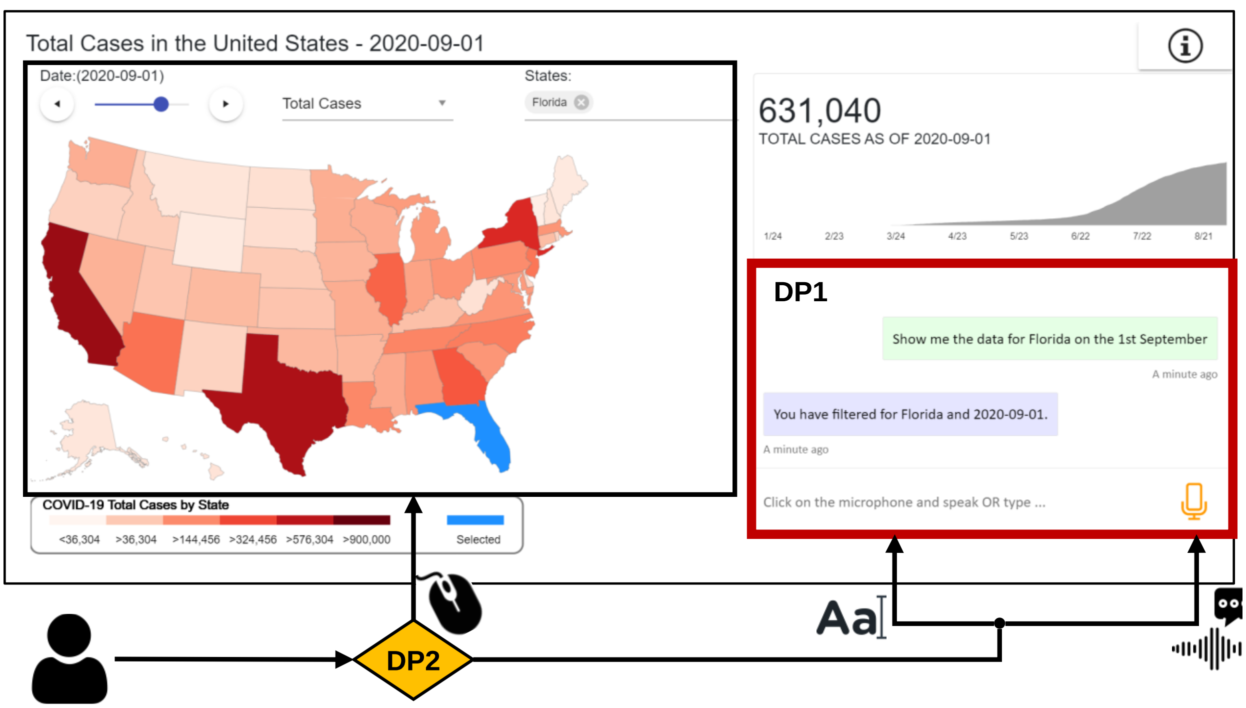

Figure 3. Screenshot of the Conversational Dashboard with DP1 and DP2

3.5 Artifact Description

To instantiate our DPs in an artifact, we developed a system architecture and implemented four key components. To ensure replicability and provide practitioners with actionable guidance on how to translate our DPs into appropriate features (Lukyanenko et al., 2020), we leveraged existing open-source frameworks and libraries rather than developing components from scratch. Next, we present a detailed description of the overall system architecture (see Figure 2), its four key components, and its conversational onboarding.

3.5.1 Dashboard and Data Visualization Component

The core component of our artifact is the conversational dashboard itself, which provides information about the COVID-19 pandemic through several data visualizations (e.g., charts, KPIs, maps) and offers users two ways of interacting with these visualizations: using natural language and a mouse. We identified common interaction types in current crisis response dashboards and decided to provide users with the ability to filter the data displayed in a visualization, to roll-up (abstract), and to drill-down (elaborate) the data at the state level. We used D3.js to create interactive data visualizations (Bostock et al., 2011) based on publicly available data from Johns Hopkins University's COVID-19 data repository (Dong et al., 2020). To change visualizations in real time, the data visualization component retrieves the required data from the database through a query module and provides it to the visualization rendering module, which then updates the data visualizations on the dashboard.

3.5.2 Interaction Management Component

The interaction management component is responsible for managing the communication between the event listeners that capture user interactions (e.g., button clicks, natural language input) and the corresponding functionality of the dashboard. For example, when a user selects a state in the drop-down menu, the event listener captures the interaction type (i.e., filtering) and the selected state so that the interaction management component can decide what dashboard functionality to invoke. In line with DP1, we connected this component to the NLP component that provides users with the ability to use spoken or written language. While natural language input in written form is directly sent to the NLP component for further processing, spoken user input is first translated into written text by the speech-to-text feature provided by Microsoft Cognitive Services. After the user input has been processed successfully, the results are returned to the interaction management component, which then adjusts the data visualizations accordingly. The mapping between the results provided by the NLP component and the dashboard functionality is implemented as a rule-based approach due to its finite nature. In line with DP2, the interaction mapper allows users to choose and switch between natural language and mouse interactions at any point in time depending on their preferences. As Figure 3 illustrates, users can set a filter for Florida, for example, either using natural language (e.g., "Show me the data for Florida") or by selecting Florida from the drop-down menu using their mouse.

3.5.3 Natural Language Processing (NLP) Component

To develop the NLP component, we used Microsoft's Bot Framework (Microsoft, 2021), a comprehensive open source framework for building conversational AI systems, which enables developers to create and manage conversation flows. In the following, we explain our implementation along the three subcomponents of (1) natural language understanding, (2) dialog management, and (3) natural language response generation (McTear et al., 2016).

Natural language understanding (NLU): In contrast to traditional mouse interaction where a click directly triggers an action on the dashboard, a user's natural language input (e.g., "Show me the data for Florida") first needs to be analyzed to identify the user's goal (e.g., filtering for Florida). For the development of the NLU subcomponent, we used Microsoft's Language Understanding and Intelligent Service (LUIS). LUIS enables developers to create and train custom, purpose-specific language models by leveraging preexisting and prebuilt language models (Sahu, 2017). Using LUIS, we created a custom language model to extract relevant entities (e.g., dates, state names, metrics) and to recognize users' intent (e.g., filter, drill-down, roll-up) from their spoken or written input. To create and train our language model, we performed two steps: First, since the language model had to be capable of extracting relevant entities, we derived an entity hierarchy with state names, dates, and metrics as entities, together with their possible values, from our database (e.g., all state names for the entity "state"). Subsequently, we integrated the entity hierarchy into LUIS to perform the entity extraction task through keyword matching (i.e., for state names, metrics) and prebuilt entities provided by LUIS (i.e., for dates). Second, the language model had to contain intents for each possible interaction type on the dashboard (i.e., filter, drill-down, roll-up), which could be mapped to users' natural language input. Thus, we created three intents with a set of training examples and identified entities that had to be included in a user input together with each intent. Since user input might not map to any of the possible interaction types, we also created the fallback intent for unspecific input such as "Hey" or "What can I do?" Table 1 provides an overview of intents, entities, and examples. Finally, we refined our language model using training data collected from 27 Amazon Mechanical Turk workers who were asked to provide different formulations for each possible interaction type on the dashboard. The final model included 23 unique training examples for the filter intent, 17 for drill-down, and 9 for roll-up.

Table 1. Intents, Entities, and Examples in the Language Model

Intent	Example	Required entities	Possible entities
Filter	*"Show me the data for **Florida** on the **1st of September,"** "Show me **deaths**"*	At least one possible entity	Metric; date; state
Drill-down	*"Go to **Texas**"*	State	Metric; date
Roll-up	*"Go to overview"*	-	-
Fallback	"Hey," "Blue," "What can I do?"	-	-

Dialog management: The dialog management subcomponent maintains the dialog state, tracks the state of the dashboard, and generates a system action based on the previously extracted intent and entities. Using Microsoft's Bot Builder SDK for .NET V4 (Microsoft, 2021), we implemented the following three key features: context handler, confirmation strategy, and error handling strategy. The context handler is primarily responsible for determining whether an action can be carried out on the dashboard based on its current state. For this, the context handler uses a rule-based approach to first check whether the entities extracted from the user's input satisfy the requirements of the recognized intent (see Table 1). Additionally, it continuously tracks the dialog and dashboard state at runtime in a local storage object. Based on the dashboard's current state, the context handler then checks whether the action type mapped to the intent is valid or whether constraints apply. If the context handler deems an action to be valid, it invokes the confirmation strategy, updates its current state, and forwards the recognized intent with the extracted entities to the interaction management component. However, if the context handler deems an action to be invalid—for example, when user input with zero entities is mapped to the filter or drill-down intent or if the fallback is triggered—it invokes the error handling strategy to inform the user that their desired action could not be performed on the dashboard.

Natural language response generation: Regardless of whether the confirmation or error handling strategy is invoked, the dashboard responds to users after they have provided input, giving explicit feedback about what actions were performed. Thus, the natural language response generation subcomponent, which is a crucial component of any CUI (McTear et al., 2016), enables turn-by-turn conversations between the dashboard and its users, consistent with our objective of designing a *conversational* dashboard. Its key feature is a rule-based feedback generator that uses predefined response templates (see Table 2) to provide informative feedback when the confirmation strategy is invoked and suggestive feedback when the error handling strategy is invoked or the fallback intent is triggered.

3.5.4 Conversational Onboarding

To instantiate DP3, we implemented step-by-step conversational onboarding through which users can familiarize themselves with using natural language to interact with the dashboard (see Figure 4). When users access the conversational dashboard for the first time, they are asked to complete the onboarding before they can start interacting with the dashboard. Following the suggestions of Gupta and Bostrom (2009), we implemented the following features in our conversational onboarding that correspond to high levels of the enactive learning dimensions (e.g., structuredness and restrictiveness of practice, feedback). To help users practice the essential skills for interacting with the dashboard using natural language, we focused their practice on how to formulate natural language input for the core dashboard functionalities such as filtering, drill-down, and roll-up. Further, we restricted the practice flow to a predefined sequence so that users are initially introduced to the basic actions with exemplary input and then gradually learn more complex actions that combine several basic ones. After each demonstration of an action on the dashboard, users are prompted to immediately reproduce it in order to minimize the lag between the demonstration and users' practice. Finally, users receive immediate feedback on their natural language input. For example, if relevant entities were missing in their natural language input, users are informed that not all entities were included in order to reproduce the action on the dashboard.

Table 2. Natural Language Response Templates

Strategy	Intent	Response template	Example
Confirmation strategy	Filter	*You have filtered for* ***State****,* ***Metric****, and* ***Date****.*	*You have filtered for Idaho, Cases, and 2020-05-15.*
	Drill-down	*You have selected* ***State*** *{for* ***Date*** *and* ***Metric****}.*	*You have selected Florida for 2020-08-29.*
	Roll-up	*You are back at the overview.*	-
Error handling strategy	Fallback	*You can use the following commands to interact with the dashboard:* • ***Filter:*** *"Show me Florida for June 1st."* • ***Zoom In:*** *"Go to New York"* • ***Back to Overview:*** *"Go to overview"*	-

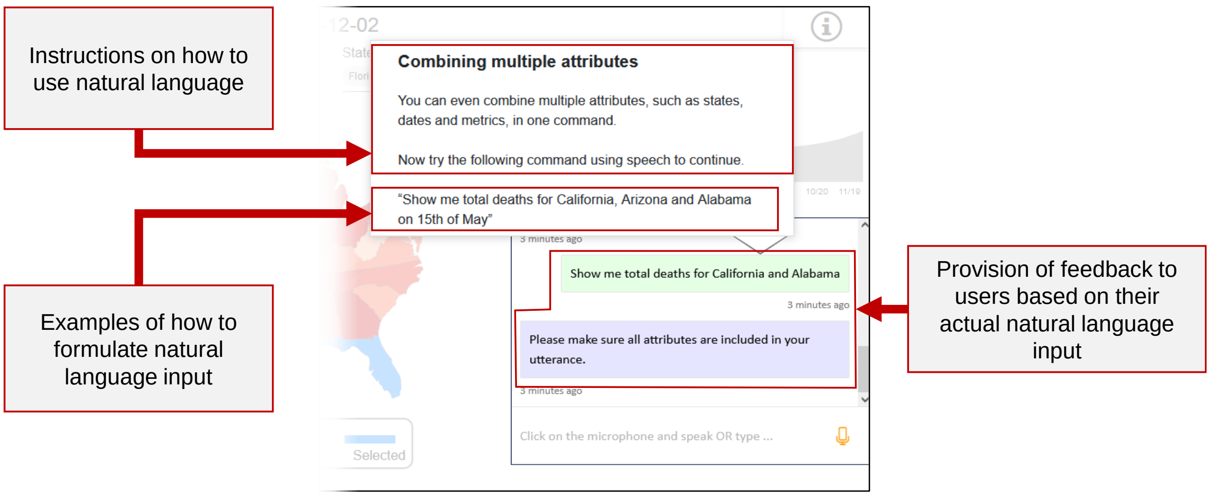

Figure 4. Screenshot of the Conversational Onboarding (DP3)

4 Evaluation

We performed two evaluations of our artifact. First, we conducted a performance evaluation of its key technical component to assess whether it enables effective natural language interaction in spoken and written form. Second, we carried out an experimental study to test whether our proposed design can improve users' transparent interaction with a crisis response dashboard.

4.1 Performance Evaluation of the Natural Language Processing (NLP) Component

At the heart of our artifact is an NLP component that allows the conversational dashboard to understand and act on the user's natural language input. To assess the NLP component's quality, we conducted a performance evaluation that specifically focused on speech-to-text translation, entity extraction, and intent mapping. For the evaluation, we used the dataset of 3119 natural language inputs collected in our user evaluation (see Section 4.2 for details). Initially, this dataset only included natural language input used to navigate the dashboard (e.g., “Show me Idaho August 31st”) and the corresponding results provided by the NLP component (e.g., intent = “filter,” entities = “Idaho” and “August 31”). Since no ground truth was available in the dataset for the evaluation, we recruited 264 crowd workers on Amazon Mechanical Turk to obtain ground truth labels for each natural language input. Additionally, we instructed workers to highlight if they recognized an input as syntactically incorrect or if there were misunderstandings (e.g., “soon out” instead of “zoom out”). Each input was labeled by two workers who had a moderate level of agreement (Cohen's kappa = 0.52). To break ties in cases of disagreement, a research assistant who received the same instructions and explanations reviewed each input with a disagreement between the workers and assigned a final label. The final dataset included 3119 natural language inputs, results of the NLP component, and human ground-truth labels for speech-to-text, entities, and intents.

4.1.1 Performance Measures

We used established measures to evaluate the performance of the NLP component. First, to verify the speech-to-text translation quality for all spoken input, we used the binary label that specified whether a particular input was syntactically correct. Based on our labeled dataset, we calculated the accuracy of speech-to-text translation as the ratio of correctly translated spoken inputs to the total number of spoken inputs. Second, to evaluate the entity extraction and intent mapping performance, we compared the results provided by our NLP component against the intent and entity labels that human workers provided. We used standard classification measures that have been used in similar work (e.g., Siering et al., 2021)—that is, precision, recall, and the F1-score—and calculated them through microaveraging the classes (i.e., intent or entity). Precision measures the percentage of correctly classified instances (i.e., intents and entities) to the total number of instances for that class of instances retrieved by the intent mapping or entity extraction ($Precision = \frac{TP}{TP+FP}$). Recall measures the percentage of correctly classified instances among all true positive cases of that class of instances ($Recall = \frac{TP}{TP+FN}$). The F1-score is calculated as the weighted average of precision and recall ($F1 = 2 * \frac{Precision*Recall}{Precision+Recall}$). Further, to benchmark our intent mapping against related systems (Srinivasan & Stasko, 2018), we additionally calculated the accuracy as the number of correctly classified inputs divided by the total number of inputs ($Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$).

4.1.2 Results

Speech-to-text: Out of a total of 3119 natural language inputs, 2499 (80.1%) were performed using spoken language. For these spoken inputs, we calculated an overall speech-to-text accuracy of 90.2%. This means that 9.8% of the spoken input included syntactical errors that could have negatively affected the subsequent entity extraction and intent mapping steps. For example, the word "cases" was incorrectly translated to "kisses" several times, resulting in user input that missed the metric "cases." However, the overall accuracy of more than 90% indicates that our speech-to-text translation was able to achieve a good performance.

Entity extraction: For the analysis of the entity extraction performance, we used all 3119 spoken and written inputs made by users, including those labeled as syntactically incorrect. Based on the F1-scores shown in Table 3, entity extraction worked the best for dates (95%) and states (94%). In contrast, entity extraction yielded a lower F1-score of 83% for the entity "metric." One reason for the lower performance of this entity is that the phrase "people had died" was used to describe "deaths," which was not included in the training data for this entity. Overall, the entity extraction step achieved an F1-score of 92%. Moreover, the results show that our entity extraction performed equally well for spoken (91%) and written input (93%). Table 3 provides the overall and entity-level precision and recall measures.

Intent mapping: The final step of the analysis was to evaluate the intent mapping performance. Again, we used all 3119 natural language inputs, including those labeled as syntactically incorrect. Only 98 inputs (3.1%) were labeled as fallback (i.e., not supported input). Overall, our intent mapping achieved a high accuracy of 82%, demonstrating comparable or better performance than related systems that offer natural language interaction with data visualizations (e.g., Srinivasan & Stasko, 2018). Additionally, the results show that our intent mapping performed equally well for spoken (83%) and written input (79%). The slight differences can be partially explained by spelling mistakes, such as "Select ketncuky," which mainly occurred in written input. In sum, our intent mapping achieved a high overall F1-score of 82% on the unbalanced dataset. While our main intents performed well, the fallback intent achieved a lower precision level (20%) because it included user input with errors from incorrect speech-to-text translation. Further, the recall for the fallback intent was only 29% since the intent mapping learned in the training phase that the phrase "Show me..." is strongly associated with the filter intent. Therefore, it also recognized inputs, such as "Show me this," as a filter intent and not as a fallback, which our artifact consequently needed to deal with since a target entity was missing. Table 3 presents all overall and intent-level results.

Response time: Finally, to evaluate the NLP component's performance in terms of speed, we analyzed the overall response time for valid inputs starting from the time a user provided spoken or written input and ending with the NLP component sending the results back to the interaction management component. The results of this analysis show that it took the NLP component only 0.9 seconds on average to fully process natural language input and update the visualizations on the dashboard accordingly.

Taken together, the results of our performance evaluation based on a dataset of 3119 manually labeled natural language inputs show that the NLP component performed well in terms of accuracy and speed, suggesting that it provides a robust technical basis to enable natural language interaction with our conversational dashboard in spoken or written form. More specifically, the NLP component achieved satisfactory performance on all tasks (i.e., speech-to-text translation, entity extraction, and intent mapping), indicating that it was able to effectively understand what users were looking for on the dashboard and to feed this information back to the other components of our conversational dashboard.

Table 3. Performance Evaluation Results for Entity Extraction and Intent Mapping

		Precision	Recall	F1	*N*
Entities	State	96%	92%	94%	2,251
	Date	96%	95%	95%	1,579
	Metric	80%	87%	83%	1,176
	Overall	**92%**	**92%**	**92%**	5,006
Intents	Filter	89%	92%	90%	2,437
	Drill-down	67%	49%	57%	472
	Roll-up	61%	54%	57%	112
	Fallback	20%	29%	24%	98
	Overall	**82%**	**82%**	**82%**	3,119

Table 4. Overview of Dashboard Types and Instantiated Design Principles

	Dashboard type	Design principles instantiated*	Description
Traditional dashboard	Traditional dashboard (*TDB*)	-	Participants were restricted to interacting with the dashboard using a mouse.
Conversational dashboard	Natural language-only conversational dashboard (*CDB-NLO*)	DP1	Participants were restricted to interacting with the dashboard using natural language.
	Natural language-enhanced conversational dashboard (*CDB-NLE*)	DP1 DP2	Participants were able to interact with the dashboard using both natural language and a mouse.

Note: *For each dashboard type, we developed two versions: one with onboarding (DP3 instantiated) and one without (DP3 not instantiated), resulting in six different dashboards used in the experiment.

4.2 User Evaluation

To evaluate whether our proposed design can improve users' transparent interaction with a crisis response dashboard, we conducted a large-scale online experiment. Following the approach of Morana et al. (2019), we developed six versions of our artifact with different combinations of instantiated DPs to examine their effect on transparent interaction. More specifically, we compared a traditional dashboard (TDB) with two types of conversational dashboards: natural language-*only* (CDB-NLO) and natural language-*enhanced* (CDB-NLE). As Table 4 shows, the CDB-NLO instantiated only DP1, whereas the CDB-NLE instantiated both DP1 and DP2. TDB did not instantiate these DPs to establish a baseline condition representing the current design of crisis response dashboards. Further, we developed two different versions of each dashboard with and without conversational and/or traditional onboarding (DP3) depending on the respective dashboard type, resulting in a total of six different dashboards.

Against the backdrop of these different artifact instantiations, we translated our previously derived design propositions (see Section 3.4) into four specific hypotheses that we empirically tested in the experiment. According to our first proposition, a crisis response dashboard equipped with a CUI should improve transparent interaction because it allows users to interact with the dashboard using spoken or written language in a natural way. Based on this proposition, we argue that users will achieve higher levels of transparent interaction with a conversational dashboard than with a traditional dashboard, regardless of whether natural language interaction replaces existing ways of interacting with a dashboard using the mouse (CDB-NLO) or whether it complements them (CDB-NLE). Hence, we hypothesize:

H1-H2: Users who interact with a natural language-only conversational dashboard (CDB-NLO; **H1**) or a natural language-enhanced conversational dashboard (CDB-NLE; **H2**) achieve higher levels of transparent interaction than those interacting with a traditional dashboard (TDB).

According to our second proposition, a conversational crisis response dashboard equipped with conversational onboarding should improve transparent interaction because it facilitates users' learning by walking them through the natural language interaction with the dashboard. Therefore, based on our second proposition, we argue that users of conversational dashboards, regardless of whether natural language interaction replaces existing ways of interacting with a dashboard using the mouse (CDB-NLO) or whether it complements them (CDB-NLE), will particularly benefit from completing the conversational onboarding before interacting with the dashboard. Hence, we hypothesize:

H3-H4: Users who complete the conversational onboarding of a natural language-only conversational dashboard (CDB-NLO; **H3**) or a natural language-enhanced conversational dashboard (CDB-NLE; **H4**) achieve higher levels of transparent interaction than those who do not.

Finally, we draw on TEU to formulate two additional hypotheses on the effects of transparent interaction on efficiency and effectiveness. TEU proposes that transparent interaction increases users' efficiency by saving them time when they navigate the system and improves their effectiveness by helping them stay focused on the task rather than getting distracted by the difficulties of finding their way around the system's interface (Burton-Jones & Grange, 2013). Based on this reasoning, we argue that higher levels of transparent interaction with a crisis response dashboard will increase users' efficiency and effectiveness in finding the information they need. Users who navigate the dashboard more quickly are able to access information in less time. In addition, they are more effective because they make fewer mistakes in their interaction and thus are less likely to give up on a task or end up with incorrect information. Hence, we hypothesize:

H5-H6: Users' transparent interaction with a crisis response dashboard increases their efficiency (**H5**) and effectiveness (**H6**) in finding the information they need.

4.2.1 Method

To test our hypotheses, we conducted an online experiment in which participants interacted with one of the six dashboards to perform four information-finding tasks. The experiment used a 3 (dashboard type: TDB vs. CDB-NLO vs. CDB-NLE) × 2 (onboarding: absent vs. present) between-subjects design, resulting in six experimental conditions.

Experimental procedure: Participants accessed the experiment via a link provided on Amazon Mechanical Turk (Mturk). After reading a short description and providing informed consent, participants were randomly assigned to one of the six experimental conditions. In the first step, participants were instructed to test their microphones to ensure that they would be able to use natural language in spoken form during the experiment and that there was only minimal background noise. Only if the system was able to understand them correctly, could they continue with the experiment. Next, participants watched a 50-second video that provided an overview of the dashboard and its COVID-19 data visualizations. After watching the video, participants in the three conditions without onboarding immediately entered the main part of the experiment. In contrast, participants in the other three conditions first completed the onboarding of their dashboard. The onboarding was designed to match the specific experimental condition so that participants only familiarized themselves with the ways of interaction that they would be able to use later. For example, the onboarding in the TDB condition did not include an introduction to natural language interaction and resembled an interactive guided tour through the GUI.

In the main part of the experiment, participants were instructed to perform four different information-finding tasks using the dashboard (see Table A2). The task order was randomized and the dashboard was reset after each task. The tasks were designed to represent realistic information needs based on our discussions with actual and potential dashboard users. For a fair comparison between different dashboard types, we designed the tasks in such a way that participants could not simply "copy and paste" the task description into the chat window and solve the task; rather, they needed to reframe it and/or break it down into multiple steps. For each task, participants could enter their solution in an input field below the dashboard or skip the task if they were unable to come up with a solution. Finally, after completing the main part of the experiment, participants filled out a survey in which they could provide feedback and report on technical problems. On average, the experiment took 25 minutes to complete.

Participants: We recruited 292 participants via mTurk. Researchers are increasingly using mTurk because the participant pool is more diverse than typical university participant pools (Buhrmester et al., 2011). Using Mturk thus supported our objective of reaching a wide range of users from different backgrounds. We excluded 21 participants who failed an attention check question, leaving 271 participants for analysis (45-46 participants per condition). Of these participants, 121 were women (44.6%) and 150 were men (55.4%). The mean age was 38.33 years (*SD* = 11.1). Table A1 shows full sample characteristics. Participants were paid \$4.5 for their participation and were able to earn a bonus payment of \$0.2 for each correctly solved task and an additional bonus of \$0.2 if they were among the 20% fastest participants for this specific task. Therefore, the maximum payment was \$6.1 (\$4.5 + 4 × \$0.2 + 4 × \$0.2).

Variables and operationalization: Transparent interaction can be assessed using self-reported measures and behavioral measures (Burton-Jones & Grange, 2013). Since self-reported measures can be subject to a range of biases and demand effects (Dimoka et al., 2011), we used a behavioral measure of transparent interaction. Following Burton-Jones and Grange's (2013) suggestions, we operationalized transparent interaction based on "the extent to which a user's navigation path … approaches the quickest path that can be taken" (p. 655). For each task in the experiment, we identified the quickest path by determining the minimum number of steps (e.g., button clicks, natural language inputs) required to navigate the dashboard to access the information needed to complete a particular task. Since this number depends on the interaction modes a dashboard offers its users (e.g., natural language and/or mouse), we calculated separate values for each dashboard type. For each participant, we then calculated the level of transparent interaction as the average ratio of the minimum number of navigation steps required for accessing the needed information to the number of navigation steps a participant actually took to correctly solve a task (see Table A3 for examples). In contrast to transparent interaction (a dimension of effective use), effectiveness and efficiency are dimensions of (task) performance. Effectiveness, which is defined as the "extent to which a user has attained the goals of the task for which the system was used" (p. 654), was operationalized as the number of correctly solved tasks. Efficiency, which is defined as "the extent of goal attainment for a given level of input (such as effort or time)" (p. 654), was calculated as the average time needed to complete all tasks that were correctly solved. Thus, effectiveness and efficiency correspond to users' higher-level goal of accessing a particular piece of information to answer a specific question (the desired end), while transparent interaction relates to users' lower-level goal of navigating the dashboard in a transparent way (the means) (cf. Burton-Jones & Grange, 2013, p. 641). Finally, we examined users' demographics (i.e., age, gender, education) and prior experience with computers, dashboards, and natural language interaction as control variables.

4.2.2 Results

Manipulation and randomization checks: We conducted two manipulation checks to ensure that participants used the different versions of the dashboard as intended. First, we asked participants to identify how they were able to interact with the dashboard (i.e., with a mouse, spoken and written language) and found that 98% of participants correctly identified their condition, which indicates that the dashboard type manipulation was successful. Second, to examine whether the onboarding successfully manipulated users' perceived ability to navigate the dashboard, we asked participants in the respective conditions before and after the onboarding to indicate their level of self-efficacy in using the dashboard on a 7-point Likert scale (Hsieh et al., 2008). The results of a paired-samples *t*-test show that participants' self-efficacy was significantly higher after completing the onboarding (M = 6.40, SD = 0.94) than before (M = 6.13, SD = 1.02; $t(134) = 3.58$, $p <$ 0.001). Moreover, participants in the conditions with onboarding rated their self-efficacy significantly higher after familiarizing themselves (M = 6.40, SD = 0.94) compared to participants in conditions without it (M = 5.90, SD = 1.04; $t(266.76) = 4.15$, $p < 0.001$). Taken together, these results suggest that the onboarding also successfully manipulated users' perceived ability to interact with the dashboard. Finally, we assessed the efficacy of our randomization procedure by comparing the six experimental conditions on several control variables. There were no significant differences in age ($F(5, 265) = 0.77$, $p = 0.57$), gender ($\chi^2(10) = 7.38$, $p =$ 0.69), education ($\chi^2(20) = 18.6$, $p = 0.55$), prior experience with computers ($F(5, 265) = 0.61$, $p = 0.69$), prior experience with dashboards ($\chi^2(20) = 15.2$, $p =$ 0.77), and prior experience with natural language interaction ($\chi^2(20) = 10.2$, $p = 0.96$). This suggests that the randomization in our experiment was also successful.

Hypothesis testing: The descriptive statistics for transparent interaction across the experimental conditions are shown in Table 5. To test our hypotheses on the effects of dashboard type and onboarding on users' transparent interaction with the dashboard (H1-H4), we conducted a two-way ANOVA. The results show significant effects of both dashboard type ($F(2, 265) = 48.3$, $p < 0.001$) and onboarding on transparent interaction ($F(1, 265) =$ 7.38, $p = 0.007$). The interaction effect was not significant ($F(2, 265) = 0.97$, $p = 0.38$). Subsequently, we used planned contrasts to test our hypotheses. First, consistent with H1, the results show that participants in the CDB-NLO condition (M = 0.60, SD = 0.25) achieved a significantly higher level of transparent interaction than participants in the TDB condition (M = 0.31, SD = 0.14; $t(265) = 9.2$, $p < 0.001$; H1 supported). However, we found no significant difference in transparent interaction between participants in the CDB-NLE (M = 0.36, SD = 0.23) and TDB condition (M = 0.31, SD = 0.14; $t(265) =$ 1.64, $p = 0.10$; H2 rejected). Further, in the CDB-NLO condition, transparent interaction shows no significant difference between participants who completed the onboarding (M = 0.63, SD = 0.24) and those who did not (M = 0.57, SD = 0.26; $t(265) = 1.31$, $p = 0.18$; H3 rejected). In contrast, in the CDB-NLE condition, participants who completed the onboarding (M = 0.42, SD = 0.25) achieved a significantly higher level of transparent interaction than those who did not (M = 0.31, SD = 0.19; $t(265) = 2.65$, $p = 0.008$; H4 supported). Overall, these results suggest that compared to traditional dashboards, conversational dashboards improve transparent interaction—particularly if participants can use natural language only and the mouse interaction option is removed (i.e., CDB-NLO). However, if users can choose between using natural language or a mouse, as in CDB-NLE, only participants who completed the onboarding achieved a higher level of transparent interaction with the dashboard. Finally, to test the remaining hypotheses on the effects of transparent interaction on efficiency (H5) and effectiveness (H6), we ran a multivariate regression with transparent interaction as the independent variable and efficiency and effectiveness as the two dependent variables. Consistent with our hypotheses and in line with TEU, the results show that transparent interaction has a significant positive effect on efficiency ($\beta = 0.58$, $p <$ 0.001; H5 supported) and effectiveness ($\beta = 0.90$, $p <$ 0.001; H6 supported).

Table 5. Descriptive Statistics for Transparent Interaction

		Onboarding	
		Absent	***Present***
Dashboard type	TDB	0.30 (0.14)	0.33 (0.15)
	CDB-NLO	0.57 (0.26)	0.63 (0.24)
	CDB-NLE	0.31 (0.19)	0.42 (0.25)
Note: Means with standard deviations in parentheses.			

Post hoc analysis: Contrary to our expectations, participants in the CDB-NLE condition did not achieve significantly higher levels of transparent interaction than participants in the TDB condition. Since participants in the CDB-NLE (vs. CDB-NLO) condition could choose whether or not to interact with the dashboard using natural language, a possible explanation could be that some of them used only the "traditional" way of interacting with the dashboard using the mouse, which might have negatively affected their level of transparent interaction. To investigate this further, we conducted a post hoc analysis of user behavior in the CDB-NLE condition. For each participant, we calculated the proportion of navigation steps they took using natural language (in both spoken and written form). This resulted in a continuous variable ranging from 0 to 1, where a value of zero indicates that natural language was not used at all. Subsequently, we ran a linear regression model with transparent interaction as the dependent variable and the proportion of navigation steps via natural language as our independent variable. The results in Figure 5 show that the proportion of navigation steps via natural language significantly influenced transparent interaction (β = 0.54, p < 0.001), suggesting that the more users interact with the dashboard using natural language, the higher their level of transparent interaction.

Since half of the participants in the CDB-NLE condition completed the conversational onboarding to familiarize themselves with how to interact with the conversational dashboard, it is conceivable that those participants also used natural language more frequently than participants who did not receive the onboarding. Therefore, we conducted a mediation analysis using the bootstrapping approach with 5000 samples (Hayes, 2017). We estimated a simple mediation model (Model 4) with onboarding as the independent variable, the proportion of navigation steps via natural language as the mediator, and transparent interaction as the dependent variable. The results show that the direct effects of both onboarding (β = 0.089, p = 0.002) and the proportion of navigation steps via natural language (β = 0.53, p < 0.001) are significant. However, the effect of onboarding on the proportion of navigation steps via natural language (p = 0.46), as well as the indirect effect of onboarding on transparent interaction through the proportion of navigation steps via natural language, are not significant (CI = [-0.03, 0.04]; p = 0.47). In summary, these results suggest that although completing the onboarding did not result in a significant increase in the use of natural language to interact with the dashboard, it helped participants to achieve higher levels of transparent interaction. A possible explanation could be that participants had learned when to choose which way of interaction and how to formulate natural language input more effectively to navigate the dashboard. The results also provide further evidence that since participants could choose between natural language and mouse interaction, some of them did not harness the potential benefits of natural language, which ultimately resulted in lower levels of transparent interaction. Put differently, on average, participants in the CDB-NLE condition did not perform better than those in the TDB condition because some of them did not leverage our new functionality but used only their mouse to interact with the dashboard. Table 6 summarizes the results of the hypothesis testing.

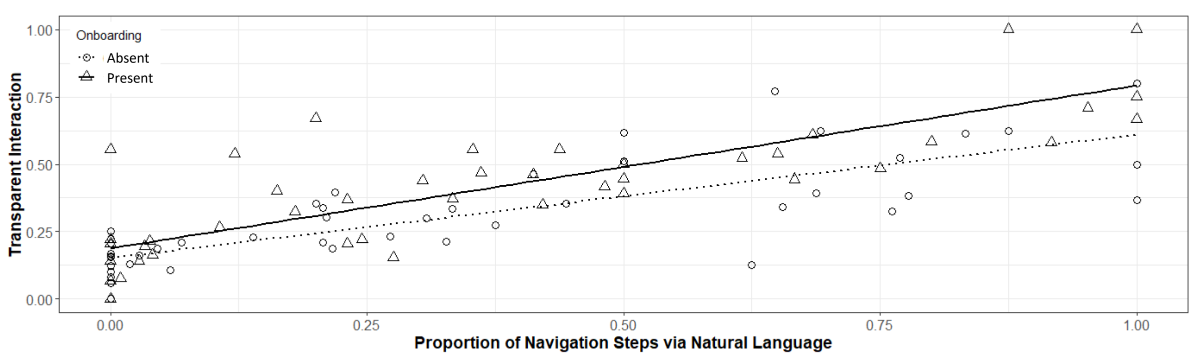

Figure 5. Effect of Proportion of Interactions Performed via Natural Language and Conversational Onboarding on Transparent Interaction (CDB-NLE)

Table 6. Summary of the Results

Hypothesis	Result	Findings
H1	Supported	Users who were able to interact with the dashboard using natural language only (*CDB-NLO*) achieved higher levels of transparent interaction than users who were able to interact using a mouse only (*TDB*).
H2	Not supported	Users who were able to interact with the dashboard using both natural language and a mouse (*CDB-NLE*) did not achieve higher levels of transparent interaction than users who were able to interact using a mouse only (*TDB*). However, a post hoc analysis showed that transparent interaction with a *CDB-NLE* depends on whether and how often users employed natural language in their interaction.
H3	Not supported	Completing the conversational onboarding of a *CDB-NLO* did not improve users' transparent interaction with the *CDB-NLO*.
H4	Supported	Completing the conversational onboarding of a *CDB-NLE* improved users' transparent interaction with the *CDB-NLE*.
H5	Supported	Transparent interaction increased efficiency.
H6	Supported	Transparent interaction increased effectiveness.

5 Discussion

Providing information to protect the public's health and safety is an important task in crisis response. In recent crises, such as the COVID-19 pandemic, many governments and health organizations developed dashboards that organize complex crisis-related data in an easy-to-digest visual format. Although these crisis response dashboards target the general public, research suggests that the average user could face difficulties in interacting with a dashboard and finding the information needed to make everyday decisions. To address this challenge, we proposed a theory-driven design for conversational dashboards intended for crisis response and developed a conversational dashboard for the COVID-19 pandemic following the DSR approach. In contrast to current crisis response dashboards, our artifact enables users to employ natural language in spoken or written form to interact with the dashboard. In addition, our artifact includes conversational onboarding that helps users familiarize themselves with how to interact with the dashboard using natural language. To rigorously evaluate our proposed design, we conducted a large-scale online experiment with six different versions of our dashboard. The evaluation results show that compared to a traditional dashboard, users achieved higher levels of transparent interaction with our dashboard, ultimately increasing their efficiency and effectiveness in finding the information they need. Moreover, the results demonstrate that the conversational onboarding supported users in learning how to interact with the dashboard, particularly when they were able to use both natural language and a mouse, which further improved their transparent interaction. Following the guidelines of Gregor and Jones (2007), we synthesized our findings into a nascent design theory for conversational dashboards intended for crisis response (see Table 7).

5.1 Theoretical Contributions

This paper makes three important theoretical contributions. Our core contribution is a nascent design theory that offers explicit prescriptions on how to extend crisis response dashboards with natural language interaction capabilities in order to improve users' transparent interaction with them and facilitate access to information. While previous research and current crisis response dashboards have focused heavily on GUIs, we propose an innovative, theory-driven design for conversational dashboards and instantiate it in a novel data science artifact: a conversational dashboard for the COVID-19 pandemic equipped with a CUI to allow natural language interaction in spoken or written form. With these findings, we contribute to research on dashboard design, both in general and in the specific context of crisis response, by delivering prescriptive knowledge for designing conversational crisis response dashboards that enable faster and easier access to important crisis-related information. More broadly, our findings also add to the data science literature by providing novel insights on how natural language can narrow the gap between the creation and consumption of insights provided by data science artifacts, particularly when they are designed for broader audiences. While prior research has emphasized key activities (e.g., data analysis, model development) in the earlier stages of the information value chain (Abbasi et al., 2016), the ultimate goal of data science is to offer actionable insights that support decision-making (Grover et al., 2018). With our focus on the latter stages of the information value chain, we complement existing data science research by providing design guidelines to help users access information on dashboards so that they can extract insights required for improved decision-making and ultimately take full advantage of such data science artifacts.

Table 7. A Nascent Design Theory for Conversational Dashboards Intended for Crisis Response

Component	Description
Purpose and scope	The purpose of the design theory is to provide prescriptive knowledge on how to design conversational dashboards intended for crisis response.
Constructs	The design theory builds on the following constructs from TEU (Burton-Jones & Grange, 2013): transparent interaction, efficiency, effectiveness, and the two drivers of effective use (i.e., adaptation and learning).
Principles of form and function	We propose three DPs for the design of conversational dashboards intended for crisis response: • DP1: *To enable the general public to seamlessly navigate a dashboard for crisis response, provide users with the ability to use spoken or written language in a natural way because articulating an information need in natural language is easier than translating it into a series of actions on the graphical user interface.* • DP2: *To enable the general public to seamlessly navigate a dashboard for crisis response, provide users with the ability to choose between natural language and mouse because it gives them flexibility for the task at hand and takes their individual preferences into account.* • DP3: *To enable the general public to seamlessly navigate a* ***conversational*** *dashboard for crisis response, provide users with conversational onboarding that takes them step-by-step through the natural language interaction with the dashboard because this helps users familiarize themselves with how to interact with the dashboard using spoken or written language.*
Justificatory knowledge	The three MRs were derived from TEU, our kernel theory. In addition, our DPs were informed by research on affordances (DP1-2) and enactive learning (DP3).
Testable propositions	We derived two testable propositions to evaluate our proposed design: • Proposition 1: *A crisis response dashboard equipped with a conversational user interface allowing users to interact with the dashboard using natural language will enable them to achieve higher levels of transparent interaction.* • Proposition 2: *A conversational crisis response dashboard equipped with conversational onboarding that walks users through natural language interaction with the dashboard will enable them to achieve higher levels of transparent interaction.*
Artifact mutability	The conversational dashboard is mutable, specifically with respect to the underlying data. While updates to the existing data can be handled without major changes, more adaptation is required for integrating new metrics (e.g., number of people vaccinated), providing new data visualizations, or supporting additional languages. With more substantive changes, the artifact could also be adapted for use in other crises (e.g., other pandemics or natural disasters).
Principles of implementation	To instantiate the DPs in our artifact, we developed a system architecture based on existing open source frameworks and libraries (see Section 3.5), which can serve as a blueprint for implementing similar artifacts.
Expository instantiation	The design theory was instantiated in an artifact: a conversational dashboard for the COVID-19 pandemic. A demonstration video can be accessed at https://youtu.be/eJZK41HDbk0.

Second, our findings shed light on potential design trade-offs that arise in providing users with multiple ways of interacting with a crisis response dashboard. As predicted, we find that users achieve higher levels of transparent interaction when they can use *only* natural language instead of *only* their mouse to interact with the dashboard, thus confirming our expectations that navigating a dashboard by articulating an information need in natural language is generally easier than translating it into a series of actions on the GUI. However, our results also suggest that when given the opportunity to use both natural language and mouse, some users prefer not to use natural language at all in interacting with the dashboard. Instead, they rely solely on the more familiar mouse interaction, which unfortunately often leads to lower levels of transparent interaction. This finding is consistent with TEU (Burton-Jones & Grange, 2013), which posits that transparent interaction is not a property of the system, but rather involves a user, system, and task. In other words, different users use the same dashboard for the same task but achieve different levels of transparent interaction because one leverages natural language while another uses the mouse. Consequently, it could be argued that natural language and mouse interaction should not be implemented together; rather, one must be chosen over the other (preferably natural language interaction). However, our results suggest that this dilemma can be addressed through conversational onboarding, which allows users not only to familiarize themselves with how to interact with the dashboard using natural language but also to learn when and where to choose which way of interacting. This might also explain why onboarding has a weaker impact on transparent interaction when users only have mouse or natural language available instead of both. In such contexts, users do not have the possibility of deciding for themselves and therefore inevitably have to deal with the benefits and challenges that come with one particular way of interacting with the dashboard. Taken together, our findings suggest that conversational onboarding is a valuable addition to conversational dashboards, even if it requires users to take an additional step before they can actually use the dashboard. In summary, these findings contribute to the emerging stream of research on novel interaction modes (e.g., Liu et al., 2021) by uncovering and addressing design trade-offs in crisis response dashboards that can be navigated using both natural language and a mouse.

Third, our research offers a methodological contribution to the IS use literature by demonstrating a novel approach for measuring transparent interaction—a key dimension of effective use—based on user interaction data. Although Burton-Jones and Grange (2013) have noted that "self-report measures alone may prove insufficient" to measure effective use objectively (p. 653), existing research has mostly relied on self-reported data (Trieu et al., 2022). Other, more objective approaches, such as the observation of users in their workplace setting (e.g., Burton-Jones & Volkoff, 2017), are often time-consuming and labor-intensive. In contrast, we use log data of user interactions with the dashboard (e.g., mouse clicks, natural language input) to provide a more objective assessment of users' level of transparent interaction by comparing their actual navigation path to the minimum number of navigation steps that are required to access a particular piece of information on the dashboard. Therefore, researchers can use our approach as a blueprint for a viable, less time-consuming alternative or supplement to existing measurement approaches to effective IS use.

5.2 Practical Implications

The outcomes of our DSR project have important implications for data science practitioners who build models, create visualizations, and develop dashboards for crisis response. Industry-standard data science processes, such as CRISP-DM (Shearer, 2000) and Microsoft's Team Data Science Process (Microsoft, 2022), emphasize that the successful deployment of data science artifacts (e.g., dashboards) and their use by the target audience is a crucial step in any data science project. The value of a data science artifact can be realized only if users are able to access it and extract insights from it (Davenport & Malone, 2021). Against this backdrop, our work can help data scientists realize the potential of natural language interaction to make their artifacts in general, and dashboards in particular, more accessible to broader audiences. To this end, the design principles, system architecture, and in-depth description of our artifact—a conversational COVID-19 dashboard—provide actionable guidance on how to leverage existing open source frameworks and cloud services (e.g., Microsoft's Bot Framework) to develop conversational dashboards that enable users to easily access information using natural language.

Additionally, our work offers practical implications for governments, health organizations, and other institutions that provide crisis response dashboards with the aim of informing the general public. Since our findings suggest that traditional dashboard designs may fail to accommodate the average user, we recommend that practitioners explore alternative ways of providing access to the information in a crisis response dashboard—for example, by using natural language. While our evaluation shows that natural language interaction could possibly replace traditional ways of interacting with a dashboard (e.g., using a mouse), some users reported that they would still prefer to have the option to revert to using their mouse or touchscreen if, for example, they were in a public space. Therefore, practitioners could first implement natural language interaction to complement rather than replace existing ways of interacting with their dashboard and, importantly, combine it with conversational onboarding to familiarize users with how and when it is best to use natural language. Following these guidelines, practitioners could make their crisis response dashboards more accessible to broader audiences and ultimately disseminate important information more effectively during a crisis.

5.3 Limitations and Future Research

Our work is subject to some limitations. First, although we provide design knowledge for a class of artifacts (i.e., crisis response dashboards), the instantiation and evaluation of DPs focus on one particular instance of this class, namely a COVID-19 dashboard. Since dashboards for other crises, such as natural disasters, might produce different kinds of data and require different data visualizations, one limitation of this DSR project is its focus on the COVID-19 pandemic. However, since many crisis response dashboards build on the same underlying technology and provide similar user interfaces, our design theory should be generalizable to dashboard implementations for other crises. More specifically, the central idea of our nascent design theory—enabling users to interact with a dashboard using natural language—is independent of the underlying data and types of visualizations in a dashboard. However, future research is needed to test our design theory in the context of other crises.

Second, our DSR project focuses on transparent interaction as one key dimension of effective use. However, the conceptualization of effective use in TEU comprises two additional dimensions—representational fidelity and informed action—that were not included in our research. Although kernel theories are rarely used as-is in DSR "due to a mismatch in terms of scope and granularity between the theoretical frameworks and the design problem" (Arazy et al., 2010, p. 457), investigating the other two dimensions of effective use in the context of crisis response dashboards would be a fruitful future research direction. Further, future work could explore other parts of TEU by, for example, providing design knowledge for physical structures (e.g., microphones, screens). Finally, there are other important challenges involving the design of these dashboards, such as their faithful representation of real-world states (Recker, 2021) and data quality (Torres & Sidorova, 2019), which also warrant further research.

Third, we used MTurk to recruit participants for our final evaluation. Although studies show that the demographics of MTurk workers are similar to that of the general U.S. population and more diverse than many other samples (Buhrmester et al., 2011), the MTurk sample might limit the generalizability of our findings. To address this limitation, we used the parameters MTurk provides to recruit participants with a wide range of sociodemographic backgrounds and experience levels (Steelman et al., 2014). However, future research should validate our findings with a nationally representative sample.

Fourth, our final evaluation was conducted in a laptop or desktop environment. Therefore, the traditional crisis response dashboard, which we compared to our conversational dashboard, only supported conventional mouse interaction. However, mobile devices, such as smartphones and tablets, might offer users additional ways of interacting with a traditional dashboard using touch (e.g., swiping, pinching). Although touch and mouse interaction exhibit similar characteristics and limitations in the context of dashboards (Srinivasan & Stasko, 2018), future research should investigate how touch interaction affects users' level of transparent interaction.

Finally, we used behavioral data to measure users' transparent interaction with the crisis response dashboard, as well as their effectiveness and efficiency in finding information. Although we followed Burton-Jones and Grange's (2013) suggestions to compare users' actual navigation steps against the "quickest navigation path" using log data, there could be other ways of calculating transparent interaction based on this data. Therefore, more research is needed to examine and compare our approach against other measurement approaches based on self-reported data.

6 Conclusion

Dashboards are important data science artifacts designed to inform the general public during a crisis. During the COVID-19 pandemic, they attracted more public attention than ever before. Although IS and HCI research have dealt with the design and use of dashboards for decades, most research has focused on dashboards for decision makers in organizations, suggesting that previous findings might not generalize well to the class of crisis response dashboards that need to be designed for broader audiences. With our research, we show how IS theories and methods can be used to improve real-world data science artifacts and, more broadly, demonstrate that the IS community in general, and DSR scholars in particular, can help the world to be better prepared for future crises.

Acknowledgments

The authors thank the special issue editors and the three anonymous reviewers for their valuable feedback and guidance throughout the review process. They also thank Jan Recker as well as the reviewers of an earlier version of this paper presented at the ECIS 2021 conference for helpful comments and suggestions. Finally, the authors thank all participants in the interviews, think-aloud sessions, and experiments.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), 1-32.

Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022). Designing personality-adaptive conversational agents for mental health care. *Information Systems Frontiers*, *1*, 1-21.

Arazy, O., Kumar, N., & Shapira, B. (2010). A theory-driven design framework for social recommender systems. *Journal of the Association for Information Systems*, *11*(9), 455-490.

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-12.

Barki, H., Titah, R., & Boffo, C. (2007). Information system use-related activity: An expanded behavioral conceptualization of individual-level information system use. *Information Systems Research*, *18*(2), 173-192.

Benke, I., Knierim, M. T., & Maedche, A. (2020). Chatbot-based emotion management for distributed teams. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1-30.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301-2309.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, *6*(1), 3-5.

Burton-Jones, A., & Grange, C. (2013). From use to effective use: A representation theory perspective. *Information Systems Research*, *24*(3), 632-658.

Burton-Jones, A., & Volkoff, O. (2017). How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. *Information Systems Research*, *28*(3), 468-489.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188.

Chen, L., Li, X., Yang, Y., Kurniawati, H., Sheng, Q. Z., Hu, H.-Y., & Huang, N. (2016). Personal health indexing based on medical examinations: A data mining approach. *Decision Support Systems*, *81*, 54-65.

Cheng, C. K. Y., Ip, D. K. M., Cowling, B. J., Ho, L. M., Leung, G. M., & Lau, E. H. Y. (2011). Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. *Journal of Medical Internet Research*, *13*(4), Article e1658.

Davenport, T., & Malone, K. (2021). Deployment as a critical business data science discipline. *Harvard Data Science Review*, *3*(1), https://doi.org/10.1162/99608f92.90814c32.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. *Business and Information Systems Engineering*, *62*(3), 193-209.

Diederich, S., Brendel, A. B., 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.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*, *22*(4), 687-702.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533-534.

Few, S. (2006). *Information dashboard design: The effective visual communication of data* (1st ed.). O'Reilly.

Flowers, J. (2020, September 4). The COVID-19 dashboard: bringing together data and statistics in one place: Public health matters. *Blog: UK Health Security*. https://publichealthmatters.blog.gov.uk/2020/09/04/the-covid-19-dashboard-bringing-together-data-and-statistics-in-one-place/

Gao, T., Dontcheva, M., Adar, E., Liu, Z., & Karahalios, K. (2015). Datatone: Managing ambiguity in natural language interfaces for data visualization. *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (pp. 489-500).

Gardner, L., Ratcliff, J., Dong, E., & Katz, A. (2021). A need for open public data standards and sharing in light of COVID-19. *The Lancet Infectious Diseases*, *21*(4), Article e80.

Gibson, J. J. (1977). The theory of affordances. In *Perceiving, Acting, and Knowing* (Vol. 1, Issue 2, pp. 67-82). Lawrence Erlbaum.

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). Opposing effects of response time in human-chatbot interaction: The moderating role of prior experience. *Business and Information Systems Engineering*, *64*(6), 773-791.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems*, *21*(6), 1622-1652.

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312-335.

Grover, V., Chiang, R. H. L., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388-423.

Gupta, S., & Bostrom, R. P. (2009). Technology-mediated learning: A comprehensive theoretical model. *Journal of the Association for Information Systems*, *10*(9), 686-714.

Gupta, S., Bostrom, R. P., & Huber, M. (2010). End-user training methods. *The Data Base for Advances in Information Systems*, *41*(4), 9-39.

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, *22*(5), 315-338.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75-105.

Hsieh, J. J. P. A., Rai, A., & Keil, M. (2008). Understanding digital inequality: Comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. *MIS Quarterly*, *32*(1), 97-126.

Huang, T.-H. (Kenneth), Chang, J. C., & Bigham, J. P. (2018). Evorus: A Crowd-powered conversational assistant built to automate itself over time. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

Ivanković, D., Barbazza, E., Bos, V., Fernandes, Ó. B., Gilmore, K. J., Jansen, T., Kara, P., Larrain, N., Lu, S., Meza-Torres, B., Mulyanto, J., Poldrugovac, M., Rotar, A., Wang, S., Willmington, C., Yang, Y., Yelgezekova, Z., Allin, S., Klazinga, N., & Kringos, D. (2021). Features constituting actionable COVID-19 dashboards: Descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. *Journal of Medical Internet Research*, *23*(2), Article e25682.

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey: Exploring the use of a context view for chatbots. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

Kim, Y.-H., Lee, B., Srinivasan, A., & Choe, E. K. (2021). Data@Hand: Fostering visual exploration of personal data on smartphones leveraging speech and touch interaction. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2021). Value co-creation in smart services: A functional affordances perspective on smart personal assistants. *Journal of the Association for Information Systems*, *22*(2), 418-458.

Koch, T. (2021). Welcome to the revolution: COVID-19 and the democratization of spatial-temporal data. *Patterns*, *2*(7), Article 100272.

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489-504.

Lauterbach, J., Mueller, B., Kahrau, F., & Maedche, A. (2020). Achieving effective use when digitalizing work: The role of representational complexity. *MIS Quarterly*, *44*(3), 1023-1048.

Lee, B., Choe, E. K., Isenberg, P., Marriott, K., & Stasko, J. (2020). Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications*, *40*(2), 82-90.

Leong, C., Pan, S. L., Ractham, P., & Kaewkitipong, L. (2015). ICT-enabled community empowerment in crisis response: Social media in Thailand flooding 2011. *Journal of the Association for Information Systems*, *16*(3), 174-212.

Liu, S. B., & Palen, L. (2010). The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science*, *37*(1), 69-90.

Liu, Y. (Alison), Shen, Y., Luo, C., & Chan, H. C. (2021). Reach out and touch: Eliciting the sense of touch through gesture-based interaction. *Journal of the Association for Information Systems*, *22*(6), 1686-1714.

Lu, Y., Chen, S., Miao, Z., Delen, D., & Gin, A. (2021). Clustering temporal disease networks to assist clinical decision support systems in visual analytics of comorbidity progression. *Decision Support Systems*, *148*, Article 113583.

Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286-5297).

Lukyanenko, R., Parsons, J., & Hovorka, D. S. (2020). Research perspectives: Design theory indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? *Journal of the Association for Information Systems*, *21*(5), 1343-1369.

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, *37*(3), 101284.

McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational interface: Talking to smart devices*. Springer International Publishing.

Microsoft. (2021). *Microsoft bot framework*. https://dev.botframework.com/

Microsoft. (2022). *The team data science process lifecycle*. https://docs.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle-deployment

Morana, S., Kroenung, J., Maedche, A., & Schacht, S. (2019). Designing process guidance systems. *Journal of the Association for Information Systems*, *20*(5), 499–535.

Nadj, M., Maedche, A., & Schieder, C. (2020). The effect of interactive analytical dashboard features on situation awareness and task performance. *Decision Support Systems*, *135*, Article 113322.

Nguyen, A., Tuunanen, T., Gardner, L., & Sheridan, D. (2021). Design principles for learning analytics information systems in higher education. *European Journal of Information Systems*, *30*(5), 541-568.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, *28*(1), 17-48.

Park, H., Bellamy, M. A., & Basole, R. C. (2016). Visual analytics for supply network management: System design and evaluation. *Decision Support Systems*, *91*, 89-102.

Patino, M. (2021). *The rise of the pandemic dashboard*. Bloomberg. https://www.bloomberg.com/news/features/2021-09-25/why-every-government-needs-a-covid-dashboard

Pietz, J., McCoy, S., & Wilck, J. H. (2020). Chasing John Snow: data analytics in the COVID-19 era. *European Journal of Information Systems*, *29*(4), 388-404.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

Prevent Epidemics. (2020). *Tracking COVID-19 in the United States: From information catastrophe to empowered communities*. https://preventepidemics.org/wp-content/uploads/2020/07/Tracking-COVID-19-in-the-United-States-Report-1.pdf

Recker, J. (2021). Improving the state-tracking ability of Corona dashboards. *European Journal of Information Systems*, *30*(5), 476-495.

Ruoff, M., & Gnewuch, U. (2021). Designing multimodal BI&A systems for co-located team interactions. *Proceedings of the 29th European Conference on Information Systems*.

Ruoff, M., Myers, B. A., & Maedche, A. (2022). ONYX: User interfaces for assisting in interactive task learning for natural language interfaces of data visualization tools. *Proceedings of the 2022 CHI Conference Extended Abstracts on Human Factors in Computing Systems*.

Sahu, A. (2017). [Cognitive services]—Enable natural language interaction with LUIS. *msdn Magazine*, 32(7), https://docs.microsoft.com/en-us/archive/msdn-magazine/2017/january/cognitive-services-enable-natural-language-interaction-with-luis

Saktheeswaran, A., Srinivasan, A., & Stasko, J. (2020). Touch? Speech? Or touch and speech? Investigating multimodal interaction for visual network exploration and analysis. *IEEE Transactions on Visualization and Computer Graphics*, *26*(6), 2168-2179.

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, *32*(3), 736-751.

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, *22*(4), 931-967.

Setlur, V., Battersby, S. E., Tory, M., Gossweiler, R., & Chang, A. X. (2016). Eviza: A natural language interface for visual analysis. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 365-377).

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 15-18.

Siering, M., Muntermann, J., & Grčar, M. (2021). Design principles for robust fraud detection: The case of stock market manipulations. *Journal of the Association for Information Systems*, *22*(1), 156-178.

Soper, D. S., Demirkan, H., & Schlicher, J. (2021). Analytics and IT in the response to COVID-19: A research framework and lessons for the future. *Journal of Decision Systems*, *31*(1-2), 7-18.

Srinivasan, A., Lee, B., Henry Riche, N., Drucker, S. M., & Hinckley, K. (2020). InChorus: Designing consistent multimodal interactions for data visualization on tablet devices. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Srinivasan, A., & Stasko, J. (2018). Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 511-521.

Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital age: Innovative alternatives to student samples. *MIS Quarterly*, *38*(2), 355-378.

Torres, R., & Sidorova, A. (2019). Reconceptualizing information quality as effective use in the context of business intelligence and analytics. *International Journal of Information Management*, *49*, 316-329.

Trieu, V.-H., Burton-Jones, A., Green, P., & Cockcroft, S. (2022). Applying and extending the theory of effective use in a business intelligence context. *MIS Quarterly*, *46*(1), 645-678.

Vallurupalli, V., & Bose, I. (2018). Business intelligence for performance measurement: A case based analysis. *Decision Support Systems*, *111*(2017), 72-85.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3506-3510).

Young, G. W., & Kitchin, R. (2020). Creating design guidelines for building city dashboards from a user’s perspectives. *International Journal of Human-Computer Studies*, *140*, Article 102429.

Young, G. W., Kitchin, R., & Naji, J. (2021). Building city dashboards for different types of users. *Journal of Urban Technology*, *28*(1-2), 289-309.

Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., & Li, T. J. J. (2022). StoryBuddy: A human-AI collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, *2*(2), 6-32.

Appendix: Additional Material for User Evaluation (Online Experiment)

Table A1. Demographic Information of Participants

		N	%
Gender	Female	121	44.6%
	Male	150	55.4%
Age	18 - 24 years	19	7%
	25 - 34 years	93	34.3%
	35 - 44 years	87	32.1%
	45 - 54 years	41	15.1%
	55+ years	31	11.5%
Education	High School	50	18.5%
	Technical, trade, or business after high school	45	16.6%
	Bachelor's degree	136	50.2%
	Master's degree	31	11.4%
	Doctoral degree or professional degree (JD, MD)	9	3.3%
Experience with dashboards	Never	31	11.4%
	1-2 times a year	46	17%
	1-2 times a month	73	26.9%
	1-2 times a week	69	25.5%
	daily	52	19.2%
Experience with conversational user interfaces	Never	25	9.2%
	1-2 times a year	20	7.4%
	1-2 times a month	42	15.5%
	1-2 times a week	82	30.3%
	daily	102	37.6%
Computer self-efficacy	*M* = 6.01 (*SD* = 1.03)		

Table A2. Experimental Tasks

1. Which of the following states had the fewest counties with more than 40,000 confirmed cases on September 29th? (Texas, California, Florida)
2. Which of the following states had the largest increase in total cases between May 15th and August 31st? (Idaho, Kentucky)
3. Which of the following regions had more cases as of October 25th? (West, Midwest)
4. How many people had died in Wisconsin, Nebraska, Idaho, and Connecticut combined by May 3, 2020?

Table A3. Calculation of Transparent Interaction based on Users' Navigation Path: Examples

Navig. steps	**Quickest Path Task 2** ***(mouse)***	**Actual navigation path taken** (P5004790 – ***TDB*** condition)	**Level of transparent interaction**
1	Filter for Idaho	Drill-down for Idaho	TI = 5/6 = 0.83
2	Filter for 2020-05-15	Filter for 2020-05-15	
3	Filter for Kentucky	Filter for 2020-08-31	
4	Filter for 2020-08-31	Zoom out	
5	Filter for Idaho	Drill-down for Kentucky	
6		Filter for 2020-05-15	
Navig. steps	**Quickest Path Task 2** **(*natural language*)**	**Actual navigation path taken** (P.9109728 ***CDB-NLO*** condition)	**Level of transparent interaction**
1	Filter for Idaho and 2020-05-15	Filter for Idaho	TI = 4/6 = 0.67
2	Filter for Kentucky	Filter for Kentucky	
3	Filter for 2020-08-31	Filter for 2020-05-15	
4	Filter for Idaho	Filter for Idaho	
5		Filter for 2020-08-31	
6		Filter for Kentucky	

About the Authors

Marcel Ruoff is a PhD student at Karlsruhe Institute of Technology, Germany. His research focuses on the intersection between information systems and human-computer interaction by designing and developing the next generation of user interfaces for end users with limited knowledge of the systems they use. In particular, he investigates how to utilize natural language to support end users in effectively using dashboards for their data exploration and analysis and how to enable these end users to effectively teach the conversational user interfaces how to perform new natural language inputs.

Ulrich Gnewuch is a postdoctoral researcher at Karlsruhe Institute of Technology, Germany. His research interests lie in the areas of artificial intelligence (AI)-based information systems and interactive business intelligence and analytics (BI&A) systems. His work has been published or is forthcoming in journals such as *Information Systems Research*, *Journal of the Association for Information Systems*, *Business & Information Systems Engineering*, *Computers in Human Behavior,* and *International Journal of Human-Computer Studies*. He serves as a board member and vice chair for teaching resources of the AIS Special Interest Group on Human-Computer Interaction.

Alexander Maedche is a professor of information systems at Karlsruhe Institute of Technology, Germany. He heads the human-centered systems lab (h-lab), conducting research at the intersection of information systems and human-computer interaction on the design of human-centered systems for better work and life. His work has been published in journals such as *MIS Quarterly, Journal of the Association for Information Systems, Computers in Human Behavior*, *International Journal of Human-Computer Studies,* and *IEEE Transactions on Software Engineering*. He is a member of the editorial boards of *MIS Quarterly*, *Journal of the Association for Information Systems,* and *Business & Information Systems Engineering*. He is a founding member of the nonprofit associations Usability & UX in Germany e.V. and Die WI e.V.

Benjamin Scheibehenne is a professor at Karlsruhe Institute of Technology, where he founded a lab on cognition and consumer behavior. The lab applies experimental methods and mathematical modeling approaches to gain a better understanding of human judgment and decision-making processes. Before his appointment at KIT, Benjamin Scheibehenne was a full professor at the University of Geneva Faculty of Economics and Management. He received his PhD in experimental psychology from the Humboldt University in Berlin and his *Habilitation* from the Faculty of Psychology at the University of Basel in Switzerland.