

Effective use of open data portals for social good: Design principles for explainable LLM-based open data assistants[☆]

Till Carlo Schelhorn^a,^{*}, Ulrich Gnewuch^b, Alexander Maedche^a

^a Karlsruhe Institute of Technology, Kaiserstraße 89, 76133 Karlsruhe, Germany

^b University of Passau, Dr.-Hans-Kapfner-Str.12, 94032 Passau, Germany

ARTICLE INFO

Keywords:

Large language models
Explanations
Open data
Government trust
Effective use
Design science research

ABSTRACT

Public access to information through open data is critical for achieving the United Nations' Sustainable Development Goal 16, which promotes accountable and inclusive institutions. However, open data portals often present barriers, such as complex user interfaces and large data catalogs, that hinder non-technical users from effectively accessing and interpreting the data. To address this problem, we conducted a two-cycle design science research project, guided by the theory of effective use, to design an explainable large language model (LLM)-based open data assistant. After focusing on transparent interaction in the initial design cycle, we evaluated our final artifact through a large-scale online experiment with 223 U.S. citizens using official Texas open data. While natural language allows for intuitive access, our findings suggest that providing reasoning steps helps citizens better interpret data and improves their effectiveness in finding information. Unexpectedly, we also find that providing a traditional data catalog alongside the conversational interface reduced representational fidelity, suggesting that direct access to raw data may overwhelm citizens. Furthermore, the results show that our design increases perceived government transparency and trust, demonstrating its broader social impact. We synthesize these findings into two design principles for explainable LLM-based open data assistants. Our research contributes to the literature by offering valuable design knowledge for explainable LLM-based systems and by adding novel insights to the discussion about democratizing access to open data. Additionally, we provide practical guidance for implementing LLM-based open data assistants that empower citizens to hold government institutions accountable.

1. Introduction

When government information is difficult to access and understand, citizens cannot hold elected leaders accountable or participate meaningfully in the democratic process. Over time, this fosters corruption, deepens social inequality, and erodes trust in government. To prevent this, Sustainable Development Goal (SDG) 16 of the United Nations calls for developing “accountable and transparent institutions at all levels” (Target 16.6) and ensuring “public access to information” (Target 16.10) [1]. In response, many public institutions—from cities and municipalities to national governments and supranational bodies such as the European Union—have created open data portals to share information with the public and to promote government transparency [2].

However, realizing the benefits of open data requires more than publishing datasets on open data portals; these portals must also be used effectively by their intended users, namely, citizens. We approach this challenge through the lens of the Theory of Effective Use (TEU).

According to TEU, effective system use depends on achieving both transparent interaction and representational fidelity [3]. Transparent interaction refers to the extent to which users can access information in an open data portal without being hindered by its physical or surface structures (e.g., its user interface). At the same time, representational fidelity concerns the extent to which the information they retrieve is complete, clear, correct, and meaningful [3]. Both dimensions play an essential role in the effective use of open data portals. First, users must be able to easily access the available data. Second, they must be able to effectively obtain faithful representations of the underlying domain (e.g., government spending). Without both, they cannot satisfy their information needs, resulting in widespread underutilization of open data [4]. Unfortunately, prior research has identified several challenges that hinder the effective use of open data portals. The target audience of these portals comprises a diverse group with different levels of technical and analytical skills, including citizens, journalists, researchers, and civil servants [5]. Yet, open data portals typically rely on traditional

[☆] This article is part of a Special issue entitled: ‘Design Science for Social Good’ published in Decision Support Systems.

^{*} Corresponding author.

E-mail addresses: till.schelhorn@kit.edu (T.C. Schelhorn), ulrich.gnewuch@uni-passau.de (U. Gnewuch), alexander.maedche@kit.edu (A. Maedche).

data catalogs to organize their vast repositories. These catalogs often feature complex user interfaces (UIs) that make it difficult for users to find and interpret relevant information. In addition, many users are overwhelmed by the sheer volume of available data, and access alone does not necessarily enable them to extract the specific insights they need [6]. As a result, the effective use of open data portals remains low, thereby undermining their role as enablers for achieving SDG 16.

To address these challenges, prior research has proposed conversational interfaces that allow users to query open data using natural language [7,8]. However, most of these open data assistants are intent-based, requiring high implementation effort and lacking scalability across datasets and contexts [9,10]. More recently, large language models (LLMs) have introduced new opportunities for flexible natural language interaction with data. Yet most studies on human-LLM interaction in data analysis focus on technical users [11–13], overlooking the unique challenges of designing LLM-based data assistants for the diverse user base of open data portals. At the same time, LLMs raise important concerns around explainability. Their black-box nature and human-like outputs make it difficult, if not impossible, for users to assess the quality and reliability of the responses [14,15]. This issue is particularly critical in the open data context, where transparency is essential. Despite this, little research has explored how explanation mechanisms should be designed to help non-technical users understand how LLM-based open data assistants arrive at their answers.

Our research addresses these gaps by investigating the design of explainable LLM-based data assistants for the effective use of open data portals. We conducted a design science research (DSR) project [16] following the framework proposed by Kuechler and Vaishnavi [17]. In a two-cycle process, we first investigated the key issues associated with open data portal usage by reviewing relevant literature, conducting two focus groups with citizens, and interviewing civil servants responsible for open data portals on the municipal level. Drawing on TEU [3] as our kernel theory, we then derived several design requirements for an explainable LLM-based open data assistant that tackles these issues. In our first design cycle, we focused on enabling transparent interaction with the open data assistant. We evaluated the initial design in a focus group with civil servants. While highlighting the potential of LLM-based open data assistants, our results indicated that explanations are crucial for verifying the assistant's answers and enabling users to understand its inner workings, thereby increasing representational fidelity. In our second design cycle, we therefore focused on refining the design of an explanation mechanism for the open data assistant, drawing on an established explanation framework [18]. Additionally, we integrated a data catalog to replicate the typical functionality of existing open data portals, allowing us to investigate the impact of displaying raw data alongside the conversational interface. In an online experiment, we evaluated our design with 223 citizens from Texas. Our results show that while natural language explanations of the assistant's reasoning significantly enhance representational fidelity, providing a data catalog unexpectedly hinders it. Beyond effective use, we demonstrate the downstream societal impact of our design. We show that higher representational fidelity positively influences perceived government transparency, which in turn significantly enhances citizens' trust in government institutions. Based on our findings, we derived two design principles for explainable LLM-based open data assistants: the principle of transparent reasoning and the principle of interface simplification. Overall, our DSR project delivers an artifact that improves citizens' access to government open data, enables them to derive data-driven insights more easily, and ultimately strengthens transparency and trust in government.

With our research, we therefore contribute to achieving the United Nations' SDG 16 by providing actionable design knowledge for LLM-based assistants on open data portals that empower citizens to hold elected leaders accountable or participate meaningfully in the democratic process. Further, we shed light on a potential tension in the design of open data portals, showing that the provision of extensive

data catalogs can overload non-technical users and hinder representational fidelity rather than improving transparency. Finally, we extend TEU by demonstrating how representational fidelity leads to broader societal impact by increasing perceived government transparency and citizen trust.

2. Theory of effective use

Burton-Jones and Grange [3] state that to obtain the maximum benefits from information systems (IS), they must be used effectively. They define effective use as “using a system in a way that helps attain the goals for using the system” [3, p. 633]. Effective use is conceptualized as a hierarchy of three dimensions: (1) transparent interaction, (2) representational fidelity, and (3) informed action. Each dimension is necessary, but not sufficient, for the higher-level dimension. Users need to have unimpeded access to a system's representations and should not be hindered by, for example, its UI. This enables them to obtain a faithful representation of the underlying domain by the system (representational fidelity), finally allowing them to act upon these faithful representations (informed action). Effective use, in turn, positively influences performance, which is conceptualized as effectiveness and efficiency [3].

Transparent interaction is defined as “the extent to which a user is accessing the system's representations unimpeded by its surface [e.g., UI] and physical structure [e.g., computer, input/output devices]” [3, p. 642]. For example, the surface structure of an open data portal is a UI that typically consists of different menus and search features to filter for relevant datasets. Some open data portals provide interactive reports or visualizations to explore the data. Others enable downloading the underlying data as .csv files or similar formats.

The second dimension, representational fidelity, is defined as “the extent to which a user is obtaining representations from the system that faithfully reflect the domain being represented” [3, p. 642]. This means that a user obtains complete, clear, correct, and meaningful information from a system. For example, in the context of open data usage, the data presented to the users should provide them with all relevant information to understand the underlying facts (domain). Representational fidelity has a positive impact on performance (mainly effectiveness) by increasing the user's understanding and certainty of the underlying domain.

3. Design science research methodology

Our research is guided by the DSR paradigm [16] and follows the DSR framework proposed by Kuechler and Vaishnavi [17]. To address the challenges related to the ineffective use of open data portals, we designed an explainable LLM-based open data assistant grounded in the TEU [3]. TEU provides a suitable kernel theory for our artifact because it is an established framework for explaining IS use and has already been applied in related contexts, such as business intelligence systems [19] and conversational dashboards for crisis response [7]. Burton-Jones and Granges [3] propose three hierarchical dimensions of effective use: transparent interaction, representational fidelity, and informed action. Drawing on these dimensions, our two-cycle DSR project focused first on transparent interaction and then on representational fidelity.

In the first design cycle, we focused on establishing the prerequisite of *transparent interaction*. During **Problem Awareness**, we identified barriers to the effective use of open data through a literature review, focus groups with citizens, and interviews with civil servants. For **Suggestion**, we derived two design requirements grounded in transparent interaction to reduce entry barriers to open data portals. In **Development**, we developed an initial prototype, leveraging existing knowledge for creating LLM-based agents [20] and building on state-of-the-art frameworks such as LangChain and OpenAI [21]. The **Evaluation** involved an exploratory focus group [22] with eight civil

servants from a medium-sized German city, including IT, statistics, and domain experts. We aimed to (1) assess transparent interaction with the artifact and (2) gather insights to guide subsequent design cycles of our DSR project. We **Concluded** our first design cycle by summarizing our results and outlining directions for the next cycle. The focus group revealed a particular interest in the transparency of the LLM-based assistant and its reasoning process. These insights highlighted the need to further refine the assistant's explanation capabilities and to focus on representational fidelity in our second design cycle.

The second cycle addressed *representational fidelity* by focusing on the assistant's explanations. In **Problem Awareness**, we reviewed explainable AI literature, noting that while explanations aid decision-making, they require careful design to avoid over-reliance or confirmation bias [23,24]. Our **Suggestion** phase builds on Gregor and Benbasat's framework [18] to derive three requirements for explanations that help users understand the domain and task, rather than justifying the assistant's output. For **Development**, we refined our artifact to include a data catalog replicating existing portals, alongside a conversational interface integrating natural-language explanations. To ensure ecological validity, we populated the assistant with high-dimensional datasets from the Texas Open Data Portal. We conducted a summative **Evaluation** with 223 Texas residents to assess the design's impact on representational fidelity. To **Conclude**, we capture our design knowledge in two principles following Gregor et al. [25].

4. First design cycle: Transparent interaction

4.1. Problem awareness

Despite widespread availability, many datasets remain underutilized [4]. Through a literature review, focus groups, and interviews, we identified three key issues regarding the actual use of open data portals. First, although portals serve diverse user groups, they rarely support this heterogeneity [6], causing users to struggle with navigating vast catalogs due to comprehensive UIs and complex filtering options (Key Issue 1). Second, users are challenged in finding data due to missing advanced search functionality [6] and inconsistent labeling [26], often requiring exact domain terminology to locate resources (Key Issue 2). Finally, effective use is hindered by the high technical and analytical skills required to process raw data [5,27]; participants noted that analyzing raw files is often not feasible, while existing visualizations go underutilized due to perceived complexity (Key Issue 3).

4.2. Suggestion

To address these identified issues, we derive several DRs that a possible solution should comply with. These DRs will later guide the development of our artifact. We rely on TEU [3] as our kernel theory. According to the identified issues, users of open data portals are hindered from exploring the available data by a combination of comprehensive UIs that complicate navigating the system and inexperience with using the system (Key Issue 1). This showcases the need for more intuitive interfaces that suit many different users. Past research has shown that natural language interaction can ease the use of user interfaces such as dashboards by increasing transparent interaction [7]. Yet, even when identifying relevant datasets, users often lack the technical and analytical skills to properly gain insights from the underlying data. The availability of raw data or complex reporting tools does not suit non-technical users (Key Issue 3). Existing solutions enable querying of open data with natural language by leveraging intent-based conversational agents [10,28–30]. They enable question answering based on the underlying data. However, intent-based conversational agents require substantial implementation effort and can only handle a finite set of user requests [10,31]. Our system should therefore leverage more comprehensive natural language capabilities that can handle a wide variety of requests and datasets. Providing users with easy access to the data should ultimately increase transparent interaction.

DR 1. The system should allow users to explore and query open data using everyday language to achieve higher levels of transparent interaction.

As our DSR project is located in the domain of open data with the aim of increasing governmental transparency and institutional accountability, our system itself should also be transparent. This aligns with the dimension of transparent interaction as users should not only have easy access to the system but also need to understand how the system generated its answer [19]. In the context of open data, users should be able to understand which data the system's answer is based on and how the data was queried. Therefore, the system should display the steps that are taken to arrive at an answer to increase transparency.

DR 2. The system should provide an overview of the analysis process to increase process transparency, enabling users to understand how their queries are processed and how results are generated.

4.3. Development

To implement our artifact addressing these DRs, we leverage LLMs as the technical foundation, as proposed in the literature on data assistants [7,31]. In contrast to traditional intent-based solutions, their extensive language comprehension and generation capabilities allow for vague user input and reduce the need for domain-specific terminology. Further, LLM-based agents can solve complex tasks like automatically querying data and data analysis [20,32,33]. Research has also shown that the reasoning steps these agents generate can be displayed to users to increase system transparency [33,34]. In the following section, we describe the technical implementation of our artifact and its features. We leveraged two Python frameworks: LangChain for implementing the LLM-based agent and Streamlit for the application's web-based conversational interface. As LLM, we relied on the OpenAI API using the GPT-3.5 Turbo, a state-of-the-art foundational model at the time of our implementation.

When using the conversational interface, the user's input is forwarded to an LLM agent in the backend. We followed the ReAct approach [20] to design the agent. It provides access to an SQL database. Four different tools enable the agent to interact with the database. The tools are functions the agent can call that (1) retrieve the five most relevant datasets for the user's input, (2) query the SQL schema for a specific dataset, (3) assess the syntactical and semantic correctness of a provided SQL query, and (4) execute a SQL query and retrieve the resulting data. Upon receiving user input, the agent prompts the underlying LLM to dynamically select the appropriate tool and input. Subsequently, based on the response from the chosen tool, the agent determines whether to provide an answer directly or to invoke another tool. This reasoning process (a sequence of thoughts, tool invocations, and observations) is returned to the frontend together with the agent's final answer. Fig. 1 shows a screenshot of the frontend of the first cycle with an exemplary user interaction.

The UI features the message history between the user and the assistant on the right and additional context to the assistant's current answer on the left. Upon sending a message, the assistant displays the different tools used to answer the question. Further details about specific actions can be accessed by expanding the individual steps, which reveal the tool inputs and return values. When data is queried from the knowledge base, both the query and results are presented below. The complete dataset is also visible, allowing users to expand the table view for more in-depth examination.

4.4. Evaluation

In collaboration with eight experts from our partner city, we organized a focus group to evaluate our prototype. Our primary objective was to gather feedback from both technical and non-technical users of the open data portal. The IT department, characterized by high technical proficiency, primarily focuses on data provisioning through the

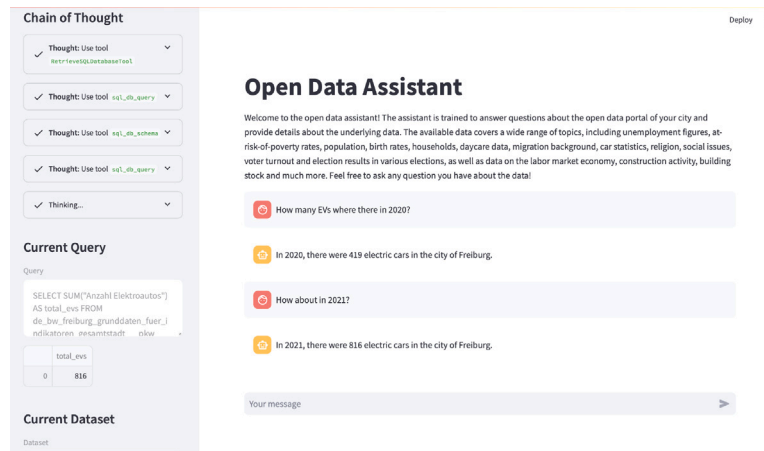


Fig. 1. Screenshot of the LLM-based open data assistant in our first design cycle showing an exemplary user interaction.

portal. Contrarily, the specialized departments represent the data consumer side, possessing significant domain knowledge but limited technical and analytical capabilities. Acting as intermediaries, the statistics department combines technical and analytical skills with domain knowledge, preparing reports, data, and responses for various city administration departments and external entities, including citizens, journalists, and researchers. Following a brief introduction, we presented our prototype to the experts, showcasing examples and explaining distinct design features. Subsequently, each expert accessed the system from their respective devices, enabling them to explore the prototype. We then gathered feedback in a discussion to understand their opinions on the prototype. In the next section, we will discuss the results of this evaluation, highlighting the identified strengths, weaknesses, opportunities, and threats.

4.4.1. Results

Our focus group evaluation results indicate an overall positive perception of our prototype. The participants acknowledged the system's usefulness for both civil servants and citizens. The evaluation revealed insights about how well the system addressed our two design requirements: enabling transparent interaction through natural language and providing an understanding of data analysis processes.

Design requirement 1: Natural language interaction. The experts praised the system's robust and everyday natural language understanding capabilities, lowering barriers to accessing open data. The system demonstrated its ability to comprehend and interpret queries, even when questions were imprecise or contained incorrect terminology. As one participant noted, "knowing the exact keyword that is used for the relevant dataset is typically necessary for finding specific data", but our system eliminated this requirement. This capability was particularly valuable for non-technical users, with another participant highlighting that many of their colleagues are unaccustomed to handling data and found that "formulating a question in natural language eases the access to the data".

The system also significantly improved the speed of information access. Participants noted that traditionally, it would take several minutes to locate the correct dataset and subsequently analyze the contained data. In contrast, they identified our system as "definitely the fastest way to extract information". However, the experts also identified some limitations in the interaction. The system occasionally misinterpreted data, presenting values from incorrect columns or performing unnecessary aggregations. While these issues did not completely impede interaction, they raised concerns about the system's reliability and the potential for misinformation.

Design requirement 2: Process transparency. The experts appreciated the system's transparency in displaying its reasoning process, which helped users understand how the system arrived at its answers. This feature enabled users to verify both the relevance of queried datasets and the accuracy of data queries. However, the current implementation of this feature received mixed feedback. While the transparency was valued, participants suggested that the technical presentation of SQL queries and data tables made the system appear more like an "expert tool", potentially unsuited to non-technical users.

The experts recommended several improvements to enhance understanding of the data analysis process. They suggested simplifying the technical aspects of the reasoning process to make the system more accessible to a broader audience. One participant recommended "supplying the relevant actions for retrieving the information directly with the answer through natural language", which would make the system's decision-making process more accessible to all users. Additionally, they emphasized the importance of including more context about the data itself. For example, when querying employment statistics, users need to understand whether the numbers include civil servants, self-employed individuals, and marginal employees in the count of socially insured people. This contextual information was considered crucial for users to properly verify and interpret the system's responses, particularly for those lacking domain-specific knowledge. Looking ahead, the experts emphasized the importance of maintaining system reliability while preserving transparency. They noted that "supplying incomplete or false information could be very harmful to users", suggesting that enabling the users to identify misinformation is essential. The system's sustainability was also discussed, with concerns raised about the environmental impact, future operational costs, and data security implications of relying on external language models.

4.5. Conclusion

The first design cycle yielded valuable insights, particularly regarding the transparency and reasoning capabilities of our LLM-based open data assistant. While the natural language interface successfully lowered barriers to accessing open data, the focus group revealed a strong interest in improving how the assistant explains its data analysis processes. These findings highlighted the need to refine the assistant's explanatory capabilities to make them accessible to non-technical users while maintaining their informativeness. This insight informed our second design cycle, which focused on enhancing the assistant's ability to explain itself in a more user-friendly manner.

5. Second design cycle: Representational fidelity

While our first cycle addressed transparent interaction by enabling navigation and exploration of open data via natural language, this dimension only pertains to accessing a system's representations. According to our kernel theory, users must also obtain faithful representations to effectively satisfy their information needs [3], captured in the second dimension of effective use: representational fidelity. In this context, users must understand the underlying data and the assistant's reasoning to leverage its answers effectively. Since our initial transparency features were considered overly technical for a significant portion of our target audience, this cycle focuses on enhancing the design of explanations. Given the extensive research on this topic, our first step was to conduct a comprehensive review of the literature on explanations and explainable AI (XAI) to understand their advantages, identify potential risks, and inform our design.

5.1. Problem awareness

Many AI models, especially LLMs, are considered black boxes due to their complexity, preventing users from understanding reasoning, verifying fidelity, or making effective decisions [3]. While research on explanations has a long history in IS [18,35], recent studies show mixed results: explanations can increase trust [36] but may not enable verification [37] or prevent confirmation bias [24]. Gajos and Mamykina [38] conclude that merely providing explanations may not encourage critical reflection, suggesting instead the presentation of information that elicits reflection on recommendations. Similarly, Dhaliwal and Benbasat [35] state that systems only improve judgments if they assist in learning or understanding the task environment. We therefore argue that special attention is needed for the explanation design of our open data assistant. Since traditional XAI methods are computationally complex for LLMs [39], we leverage intermediate steps to explain the model's reasoning process [33].

5.2. Suggestion

Against this backdrop, we draw on the established framework of Gregor and Benbasat on explanations [18] to inform the design of explanations for our open data assistant. They define three different dimensions of explanations: content type, presentation format, and provisioning strategy.

As previously discussed, different content of explanations can have different effects. Many existing approaches focus on explanations of the underlying model, providing evidence for the correctness of an output. While these justificatory explanations can increase trust in AI, it is unclear if they align with human sensemaking. They can, however, strengthen an AI model's persuasiveness, increasing overreliance, which often suppresses user reflection on its output. Schmitt [40] therefore calls for explanations that focus on the task and domain at hand rather than on the model itself. In the context of our open data assistant, rather than justifying why an answer is correct, the assistant can outline the steps taken to arrive at the answer and explain the reasoning behind each step. This enables the user to reflect on potential solutions to the task. Additionally, the assistant can describe the underlying data, helping the user assess whether the desired information is present and the steps the assistant took to retrieve it. In line with the theory of effective use, these explanations should increase representational fidelity by enabling the user to evaluate whether the assistant's analysis steps actually address the given task and whether the provided information faithfully reflects the underlying real-world domain.

DR 2.1. The system should supplement its answers with explanations about its line of reasoning to increase representational fidelity by supporting the user in better understanding the structure of available data and how it is processed.

As open data portals serve a diverse user base, explanations should be presented in a format that a broad audience can easily understand. In Section 4.2, we highlighted that natural language increases transparent interaction. Using natural language explanations ensures that the information the system provides is easy to interpret and becomes more accessible to a wide range of users, regardless of their technical expertise.

DR 2.2. The system should present its explanations in a natural language format to make them accessible to a broad audience.

How explanations are delivered is critical in ensuring their effectiveness. Users who infrequently use the open data assistant may not be aware that explanations are present if they are not displayed automatically. However, too many details at once may overwhelm users. The assistant should, therefore, automatically display its general line of reasoning for arriving at an answer. Additionally, users should be able to access additional information about the assistant's reasoning manually, if necessary. This should ensure that users recognize the availability of explanations without being overwhelmed by excessive detail.

DR 2.3. The system should provide a mix of automatic and user-invoked explanations to avoid overwhelming the user.

5.3. Development

We adapted our initial prototype to meet our newly derived DRs. While the overall architecture remained similar to that of the first cycle, we upgraded the technology stack to align with the current state-of-the-art and implemented new frontend features to support the required explanation capabilities. First, we used OpenAI's GPT 5-mini model—the latest available at the time—to power the agent. Second, instead of Streamlit, we used Vue.js, a JavaScript framework,¹ to implement a more flexible UI for our open data assistant that dynamically renders explanations. Communication between UI and agent was handled via FastAPI endpoints. Fig. 2 depicts a conceptual overview of the system architecture. In the following, we elaborate on the three main components of our open data assistant: the underlying **datasets**, the backend implementation of the **LLM agent**, and the frontend **UI**.

5.3.1. Datasets

To demonstrate the generalizability of our artifact and ensure instantiation validity [41], we decided to expand beyond the scope of our first design cycle. While the first cycle focused on a medium-sized German municipality, the artifact instantiation in the second cycle used data from a U.S. state (Texas). This shift serves to verify that our design is both dataset- and language-agnostic and to test its robustness using larger and more granular data.

In total, we sourced the 100 most-viewed datasets from the Texas Open Data Portal² using the Socrata API. To ensure system responsiveness and avoid potential context-window limitations, we excluded datasets with more than 500,000 rows, resulting in a final corpus of 89 datasets. The data pipeline involved two distinct storage mechanisms: (1) **Structured Data Storage**, where the raw content of each dataset was ingested into an SQLite database with each dataset represented as a distinct table; and (2) **Vectorized Metadata Storage**, where we aggregated the metadata (title, description, category, tags, and column schema) of each dataset into a single document to enable semantic search over exact keyword matches. We then generated vector embeddings for these documents using the *text-embedding-ada-002* model and stored them in a FAISS vector index [42].

¹ <https://vuejs.org>

² <https://data.texas.gov>

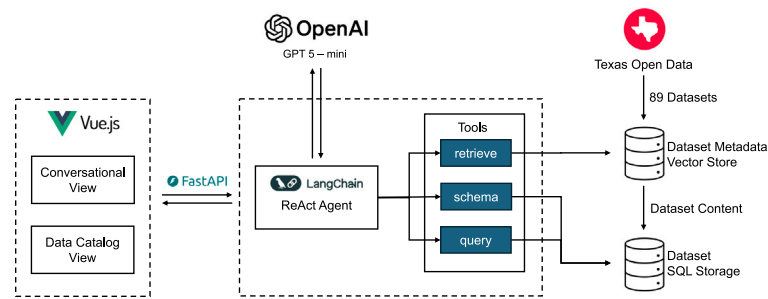


Fig. 2. Conceptual architecture of the final artifact.

5.3.2. LLM agent

The backend logic utilized LangChain’s SQL agent framework³ to interact with the SQLite database and the FAISS vector index, where all datasets are stored. This architecture enables the LLM agent to autonomously select and execute different tools (functions) necessary to answer a user’s request. For example, the agent might first find all relevant datasets to a user’s question, then retrieve the schema of the most promising dataset, and finally query data to answer the question. To this end, the agent had access to three tools:

The Retrieve step performs a semantic similarity search on the FAISS index using a natural language search term derived from the user request (Input). This process identifies the top-5 most relevant datasets, outputting their metadata (title, description, columns) (Output). Next, the Schema step extracts the structural definition for a list of target dataset names (Input), providing the SQL ‘CREATE TABLE’ statements that describe the table schema as the Output. Finally, the Query step executes an SQL query (Input) against the SQLite database, yielding the query execution results (rows of data) as the Output.

Upon receiving a user request, the LLM agent is prompted to stream both its final response and the intermediate “thought” traces (generated using the tools described above) to the frontend for display in the user interface.

5.3.3. User interface

The design of the UI followed a hybrid approach with two distinct views: (1) the **Chat View**, facilitating conversational interaction and explanation display, and (2) the **Data Catalog View**, which mimics open data portal functionality by providing access to all retrieved datasets during user interaction (see Fig. 3).

Chat view and explanations. The Chat View enables standard natural-language interaction via text messages. However, to address DR2.1 (Explain Reasoning), we integrated an explanation feature that dynamically visualizes the agent’s decision-making process. Instead of only providing the final response, the UI renders the agent’s intermediate “thought” traces as structured steps before the final answer (see Fig. 3).

To meet DR2.3 (Mix of Automatic/User Invoked), we implemented an expandable design strategy. The system automatically displays high-level information about each step in the agent’s processing of the user request, but the detailed output is hidden by default to avoid overwhelming the user. Users can request more detailed natural language explanations (DR2.2) by expanding specific reasoning steps. The Retrieve (1a) step displays the search terms used and lists the tables identified. Schema (1b) shows which table structures were inspected and provides justification for the selection. Query (1c) provides a natural language description of the executed SQL query and displays the result returned by the database.

Data catalog view. Complementing the chat view, we implemented a Data Catalog View to faithfully represent the status quo of current open data portals, providing users with direct raw access to the underlying tables. It lists all datasets identified based on the user’s requests during the session as expandable tiles (2a). Users can preview the dataset description and an excerpt of 5 columns and 5 rows (2b) or maximize a dataset to enter a “Table View” (Fig. 4). This view supports manual exploration, allowing users to filter, sort, and inspect a sample of up to 1000 rows.

5.3.4. Technical performance evaluation of the open data assistant

To ensure technical robustness, we conducted a technical performance evaluation using a set of 15 user requests generated by the GPT-5.1 model. These requests covered the three experimental tasks (five per task) of our main evaluation, using varied linguistic styles, and included context-dependent follow-up questions for Task 3. Each request was executed 10 times, resulting in a total of 150 interactions. We instructed a research assistant to manually label the responses and found that the assistant correctly answered 80% of the time. The observed errors were primarily due to incorrect SQL aggregation (e.g., using MAX instead of SUM), missing filters, magnitude errors (e.g., millions instead of billions), or failures to locate specific tables or columns. Despite these occasional errors, the results confirm that the assistant is sufficiently reliable for handling citizens’ requests regarding open data.

5.4. Evaluation

We conducted a summative evaluation of the data assistant through a controlled online experiment with U.S. citizens from Texas. The primary objective was to examine the impact of our design on users’ ability to achieve representational fidelity. Additionally, we explored its downstream effects on perceived government transparency and trust to assess the potential broader societal benefits of our design. In the following, we first derive three testable hypotheses and then describe our study design and results.

5.4.1. Testable design hypotheses

Our kernel theory, TEU, posits that *representational fidelity*—the extent to which a user obtains complete, clear, correct, and meaningful output from a system—is a prerequisite for effective system use [3]. AI-based systems often operate as “black boxes”, leaving users in the dark about how they generate their outputs and arrive at their answers. We argue that by exposing the internal reasoning process (the “how”), users can better assess the validity and faithfulness of the assistant’s output. Therefore, we propose that:

³ <https://docs.langchain.com/oss/python/langchain/sql-agent>

The figure shows a chat interface on the left and a data catalog on the right. The chat interface has three numbered steps: 1 (Which agency had the most spending in 2024?), 1a (Identified 5 relevant datasets for these search terms: most spending agency in 2024), 1b (Selected the following datasets for analysis and gathered details: texas_state_expenditures_by_county_2024), and 1c (Analyzed the selected dataset to generate the answer: The agency with the most spending in 2024 based on the amount column provided in the dataset. Result (first 5 Rows): HEALTH AND HUMAN SERVICES COMMISSION \$2,659,367,652.84). The data catalog on the right shows a list of datasets (2a) and an expanded view of the 'texas_state_expenditures_by_county_2024' dataset (2b) with a table of spending data.

Fig. 3. Exemplary user interaction in the Chat View showing the explanations for each reasoning step: (1a) dataset retrieval, (1b) schema retrieval, and (1c) data querying. In addition, the Data Catalog View shows an expandable tile for each dataset (2a) and, upon expansion, a dataset preview and its description (2b).

The figure shows a data catalog table with columns: fiscal_year, agency_number_and_name, county, major_spending_category, amount, agency_number, and agency_name. The table is maximized to show 1000 rows (3a). The interface includes search filters (3b) and sorting options (3c).

Fig. 4. Maximizing the table shows a more detailed preview of 1000 exemplary rows of the table (3a) and enables the user to search through the whole dataset (3b) or filter and sort individual rows (3c).

H1. Users achieve higher levels of representational fidelity when interacting with an LLM-based open data assistant that explains its internal reasoning process compared to when this process is opaque.

Current open data portals often feature comprehensive data catalogs to help users discover, explore, and access available datasets [8]. As a result, these data catalogs should increase users' ability to achieve representational fidelity by allowing them to understand how datasets relevant to their questions are structured. However, the literature suggests that the complexity of these data catalogs may overwhelm non-technical users [6]. Observations from our focus group sessions in the first design cycle also indicate that data catalogs may present information in ways that are not always easy for citizens to interpret. Nonetheless, removing the data catalog view from our assistant could diminish transparency, as users would no longer be able to see which datasets were retrieved in response to their requests and explore how these datasets are structured. Taking both considerations into account, we propose that:

H2. Users achieve lower levels of representational fidelity when interacting with an LLM-based open data assistant that is restricted to natural language interaction compared to one that additionally supports manual data exploration.

According to TEU, representational fidelity primarily improves effectiveness by improving users' understanding of the domain being represented and reducing their uncertainty about it [3, p. 644]. For example, when users understand what data is available and how the

data assistant processed it, they are more likely to find the information that satisfies their information need. Therefore, we propose that:

H3. Higher levels of representational fidelity during the interaction with an LLM-based open data assistant increase users' effectiveness in finding the information they need.

5.4.2. Experimental design

To test our hypotheses, we employed a 2×2 between-subjects factorial design. The experimental manipulations focused on two design dimensions: **System Explainability** (Opaque vs. Explainable) and **Interface Modality** (Hybrid vs. Conversational). Consequently, participants were randomly assigned to one of four experimental conditions. A **Hybrid Data Assistant (HDA)** served as our baseline condition featuring a hybrid approach to the interface design (i.e., Chat + Data Catalog) with an opaque assistant. We chose this as our baseline because it reflects the most plausible evolution of open data portals: retaining the traditional data catalog while augmenting it with a conversational interface. The **Explainable Hybrid Data Assistant (xHDA)** supplemented the hybrid modality with the explanation features (Reasoning steps). In contrast, the **Conversational Data Assistant (CDA)** utilized the Conversational modality by removing the data catalog to focus interactions solely on the chat interface. Finally, the **Explainable Conversational Data Assistant (xCDA)** combined the chat-only interface with the assistant's reasoning explanation capabilities. Throughout the study, we collected both behavioral metrics (clickstream data) and

survey responses to assess participants' perceptions of the assistant and their effectiveness in finding the answers to the experimental tasks.

5.4.3. Participants

We recruited participants via the CloudResearch Connect platform. To ensure high ecological validity, we restricted the sample to U.S. citizens currently residing in Texas. This geographic targeting was chosen for three primary reasons. First, it ensured that the experimental tasks, which involved investigating state agency expenditures and local county spending, were more likely to be personally relevant to the participants, thereby creating a realistic scenario of citizens holding institutions accountable. Second, a comparative analysis of state open data portals revealed that the Texas portal offers a high volume of granular datasets suitable for complex queries. Third, targeting a U.S. state with a large population ensured a sufficiently broad participant pool to facilitate data collection.

A total of 235 participants initially completed the study. We applied rigorous data quality screening, excluding 12 participants who either failed attention checks or did not interact with the open data assistant at all. This resulted in a final sample of 223 valid responses, which were roughly evenly distributed across the four conditions.

5.4.4. Experimental procedure and tasks

Upon recruitment via CloudResearch, participants were redirected to our experimental environment implemented using oTree [43]. The procedure began with onboarding, during which participants received a brief overview of the study context and incentive structure, followed by informed consent. Next, participants were randomly assigned to one of the four experimental conditions. Before beginning the tasks, we collected demographic data and introduced the assistant's specific functionality (depending on the assigned condition) through a series of annotated screenshots. During the subsequent task execution phase, participants were presented with three tasks of increasing difficulty, each requiring the retrieval of specific information regarding Texas state expenditures: Task 1 asked participants to identify the government agency with the highest fund usage in 2024 and the exact amount spent; Task 2 instructed participants to select the county they live in and identify and rank the top three spending categories for 2024; and Task 3 required participants to analyze the Texas Senate's travel expenses by comparing spending between the 2023 and 2024 fiscal years. The tasks were designed such that they could not be solved through simple web searches or general-purpose AI assistants like ChatGPT or Gemini because they required filtering and aggregating data from the state's open data portal. Finally, participants completed a post-test questionnaire capturing the dependent variables for the constructs of TEU and social impact (i.e., government transparency and trust).

To ensure high engagement, we employed a performance-based incentive structure. Participants received a base payment of \$6 for completing the study, which took on average 21.23 ($SD = 12.88$) minutes. Additionally, a bonus of \$1 was awarded for every correctly solved task, allowing for a maximum total compensation of \$9.

5.4.5. Measurements

We assessed the study constructs using established survey and behavioral measures. All items were scored on 7-point Likert scales ranging from "Strongly Disagree" (1) to "Strongly Agree" (7). Representational fidelity was measured using four items adapted from Burton-Jones and Grange [3] and Trieu et al. [19]. Users' perceived transparency of the Texas government, enabled through the open data assistant, was captured using items adapted from prior conceptualizations of governmental openness and accountability [5,44]. Trust in government was measured with items adapted from the trust framework of [45]. We confirmed the reliability and validity of our measures. Users' effectiveness in finding information to satisfy their information need was operationalized as the percentage of correctly answered questions of the experimental tasks. In addition, we logged users' clicking behavior on the assistant's reasoning steps and data catalog. Table 1 provides an overview of all measures.

5.4.6. Manipulation and randomization checks

To verify the effectiveness of our experimental manipulations, we compared participant responses to two manipulation-check items (measured on a 7-point Likert scale) using one-way ANOVAs. First, to assess the **System Explainability** manipulation, participants responded to the item: "The open data assistant provided an explanation in the chat describing how it arrived at its responses". The analysis revealed a significant effect of condition ($F(1,221) = 57.95, p < .001, \eta_p^2 = .208$), confirming that participants in the reasoning conditions ($M = 5.27, SD = 1.58$) perceived the explanations as intended compared to those in the opaque conditions ($M = 3.47, SD = 1.94$). Second, to assess the **Interface Modality** manipulation, participants responded to: "The open data assistant displayed a list of identified datasets on the right and allowed me to explore them on my own". The analysis also showed a significant effect of condition ($F(1,221) = 84.78, p < .001, \eta_p^2 = .277$), confirming the successful manipulation of the data catalog's presence, with participants in the hybrid conditions reporting higher agreement with the statement ($M = 6.13, SD = 1.02$) than those in the conversational conditions ($M = 4.17, SD = 2.04$).

To assess the effectiveness of our randomization procedure, we examined whether participants differed on key demographic and background variables across conditions. A series of one-way ANOVAs indicated no significant differences in age ($p > 0.79$), AI usage ($p > 0.06$), analysis tool usage ($p > 0.91$), or SQL knowledge ($p > 0.49$). Chi-square tests of independence confirmed that participants were equally distributed across conditions regarding gender ($\chi^2(6) = 3.14, p = .79$) and education level ($\chi^2(12) = 15.89, p = .20$). Collectively, these results suggest that the randomization in our experiment was successful.

5.4.7. Hypotheses testing

To empirically test the relationships proposed in our hypotheses, we used structural equation modeling (SEM) using the lavaan package in R. Additionally, we performed an SEM-based mediation analysis to test the indirect effects and examine the full causal chain in our research model. For these analyses, we specified a structural model with two independent variables: **System Explainability** (0 = Opaque, 1 = Explainable) and **Interface Modality** (0 = Hybrid, 1 = Conversational). Our two dependent variables were representational fidelity and effectiveness. Additionally, to control for participants' technical background, we included *AI usage*, *analysis tool usage*, and *SQL knowledge* as covariates.

We first examined the effects of our design manipulations on representational fidelity (H1 and H2). The SEM results, as shown in Fig. 5, revealed a significant positive effect of system explainability on representational fidelity ($\beta = 0.204, SE = 0.097, p = .035$). Thus, **H1 is supported**. Regarding interface modality, we hypothesized that removing the data catalog from the interface would reduce representational fidelity (H2). However, contrary to our expectations, our analysis reveals a significant positive effect of the Conversational modality on representational fidelity ($\beta = 0.325, SE = 0.092, p < .001$). This suggests that the simplified, chat-only interface actually helped users achieve higher fidelity compared to the hybrid baseline. Consequently, **H2 is not supported, but the opposite effect is found**. The interaction between the System Explainability and Interface Modality conditions was not significant ($\beta = -0.148, p = .178$), indicating that combining the two conditions does not yield a synergistic effect.

Subsequently, we assessed the impact of representational fidelity on users' effectiveness in finding the information they need (H3). Consistent with our theorizing based on TEU, the SEM results show a significant positive effect of representational fidelity on effectiveness ($\beta = 0.207, SE = 0.076, p = .006$). Therefore, **H3 is supported**.

Finally, to examine the causal chain and formally test indirect effects, we followed Zhao et al. [46] and performed a SEM-based mediation analysis using a bias-corrected bootstrapping procedure with 5000 samples [47]. We used the lavaan package in R with the same control variables. In addition to H1-3, the model included the direct,

Table 1
Descriptive statistics.

	Construct	Baseline (HDA) (n = 59)	xHDA (n = 58)	CDA (n = 51)	xCDA (n = 55)	Total (N = 223)
Effective Use	Repr. Fidelity	4.46 (1.58)	5.03 (1.49)	5.39 (1.24)	5.47 (1.05)	5.07 (1.41)
	Effectiveness	0.51 (0.26)	0.54 (0.24)	0.62 (0.24)	0.62 (0.22)	0.57 (0.25)
Societal Impact	Gov. Transp.	4.27 (1.28)	4.38 (1.39)	4.43 (1.28)	4.76 (1.32)	4.48 (1.40)
	Trust in Gov.	3.68 (1.68)	3.57 (1.73)	3.78 (1.77)	4.10 (1.70)	3.78 (1.72)

Note: Means with standard deviations in parentheses.

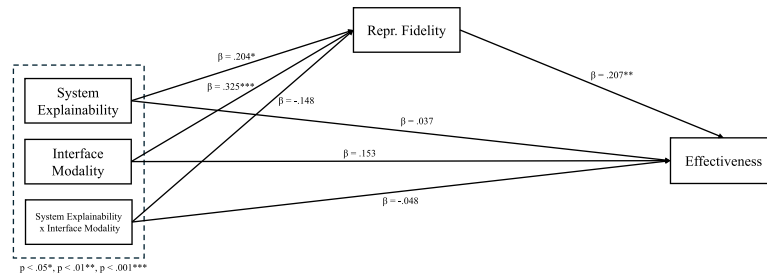


Fig. 5. Results of the structural model analysis.

indirect, and total effects of both experimental treatments and their interaction on effectiveness.

The analysis of indirect effects confirms this mechanism: the indirect path from System Explainability on effectiveness via representational fidelity was significant, as indicated by a bootstrap confidence interval ($\beta_{indirect} = 0.042$, 95% CI [0.000, 0.104]). Similarly, the indirect path of Interface Modality to effectiveness via representational fidelity was also significant ($\beta_{indirect} = 0.067$, 95% CI [0.013, 0.142]). The analysis confirmed that the direct effects of the design features, System Explainability ($\beta = 0.037$, $p = .697$, 95% CI [-0.151, 0.220]) and Interface Modality ($\beta = 0.153$, $p = .117$, 95% CI [-0.041, 0.341]), on effectiveness were not significant, suggesting full mediation. These findings suggest that the assistant's design features facilitate effective use primarily by enhancing the user's level of representational fidelity.

5.4.8. Post Hoc analysis of user interaction with the data catalog

To investigate the unexpected finding regarding H2, where the *Conversational* modality (Chat-only) yielded higher representational fidelity than the *Hybrid* modality (Chat + Data Catalog), we conducted a post hoc analysis on participants' behavioral data. We specifically used the behavioral logs of participants in the **Explainable Hybrid (xHDA)** (Reasoning Steps + Data Catalog) condition ($n = 58$). We modeled representational fidelity as the dependent variable in a multiple regression analysis. The independent variables were the log-transformed interaction counts: *Table Clicks* (interaction with the Data Catalog) and *Reasoning Clicks* (interaction with the explanation steps). As in previous models, we controlled for *AI usage*, *analysis tool usage*, and *SQL knowledge*. The regression results, visualized in Fig. 6, revealed a significant negative relationship between interaction with the data catalog and representational fidelity ($\beta = -0.427$, $SE = 0.212$, $p = .049$). This indicates that the more participants clicked on and explored the raw data tables, the lower their level of representational fidelity. In contrast, interaction with the reasoning steps did not significantly affect representational fidelity ($\beta = -0.249$, $p = .351$).

These findings offer a plausible explanation for the rejection of H2. Rather than validating the system's output, the data catalog likely introduced information overload or complexity, leading to confusion among users. Consequently, users who were shielded from this complexity (the *Conversational* condition) or those who interacted less with the catalog reported higher levels of representational fidelity.

5.4.9. Supplementary analysis: From effective use to societal impact

While our evaluation primarily focused on the effective use of the data assistant (representational fidelity and effectiveness), the ultimate goal of open data initiatives is to foster broader societal outcomes, such as making government actions transparent and increasing citizens' trust in public institutions. To assess whether our design principles translate into these downstream benefits, we conducted a supplementary analysis.

Open data initiatives aim to foster government transparency. However, transparency is not merely the publication of data, but also the capacity of citizens to access, use, and make sense of it [48]. If users cannot make sense of the data (low representational fidelity), the mechanism for transparency fails. Conversely, a faithful representation should enhance the perception that the government is open and accountable. Finally, literature suggests that perceived transparency is a critical antecedent to institutional trust [5]. If the assistant enables users to perceive the government as transparent and accountable, this should, in theory, translate into higher trust in the government.

To empirically examine these relationships, we conducted another SEM-based analysis to test the relationships between *representational fidelity*, *perceived government transparency*, and *trust in government*. We specified a model with representational fidelity as the independent variable, transparency as the mediator, and trust as the dependent variable. Similar to our main analysis, we controlled for participants' *AI usage*, *analysis tool usage*, and *SQL knowledge*. The results provide strong support for the proposed relationships. First, we found a significant positive effect of representational fidelity on perceived government transparency ($\beta = 0.409$, $SE = 0.063$, $p < .001$). Second, the analysis also reveals a significant positive effect of government transparency on trust in government ($\beta = 0.699$, $SE = 0.065$, $p < .001$). Finally, we analyzed the mediating role of transparency in the relationship between representational fidelity and trust. The results from a SEM-based mediation analysis using the bias-corrected bootstrapping procedure with 5000 samples confirmed that the indirect effect of representational fidelity on trust via government transparency was significant and positive ($\beta_{indirect} = 0.286$, 95% CI [0.197, 0.385]). The direct effect of representational fidelity on trust in government was not significant ($\beta = -0.010$, $p = .852$), indicating full mediation. This pattern of results suggests that representational fidelity does not directly increase trust; rather, it fosters a perception of transparency, which in turn significantly enhances users' trust in the government.

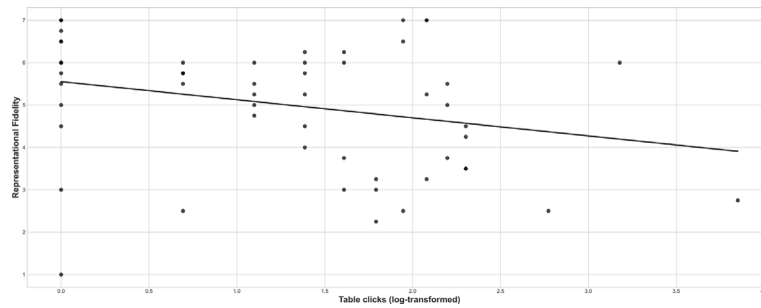


Fig. 6. Regression analysis predicting representational fidelity based on data catalog interactions in the Explainable Hybrid (xHDA) condition.

6. Discussion

Open data is crucial for fostering transparency in government institutions, improving public services, and driving innovation. Open data portals play a significant role in democratizing access to this data, enabling citizens and other stakeholders to access information about their institutions and make data-driven, informed decisions. These portals are therefore instrumental in achieving the UN’s SDGs, particularly Goal 16, which highlights the need to “[...] build effective, accountable and inclusive institutions at all levels” and “[...] public access to information”. However, many open data portals remain underutilized due to a range of barriers that hinder citizens from effectively accessing and interpreting the data [6].

To overcome these barriers, we conducted a two-cycle DSR project to design and evaluate an explainable LLM-based open data assistant. To inform our design, we drew upon the TEU by Burton-Jones and Grange [3]. In the first design cycle, we established that natural language interaction is a prerequisite for transparent interaction, allowing users to access data intuitively. The second cycle focused on representational fidelity, investigating how explanations need to be designed to make the assistant’s reasoning process more transparent and comprehensible.

We evaluated our final artifact in a large-scale online experiment with 223 citizens from the U.S. state of Texas. The results show that natural language explanations of the assistant’s reasoning (e.g., detailing table selection and SQL query formulation) helped users achieve significantly higher levels of representational fidelity and ultimately improved their effectiveness in finding answers to their questions, confirming the value of our design. However, in contrast to our expectations, combining the assistant’s conversational UI with a traditional data catalog did not improve representational fidelity. Instead, a simplified interface without it led to higher levels of representational fidelity and improved effectiveness. Interestingly, increased interactions with the data catalog were significantly and negatively correlated with representational fidelity, suggesting that the data catalog might have led to confusion rather than helping users make sense of the underlying data. Finally, we found that the benefits of our design extend beyond the direct interaction with the data assistant. The results of our supplementary analysis show that representational fidelity increases perceived government transparency, which, in turn, positively affects users’ trust in government. Based on our findings, we used an inductive approach to formalize the design knowledge resulting from our DSR project. Specifically, we derived two design principles that are inherent in the design of our artifact and have been tested in our evaluation. In the following, we present these design principles and subsequently discuss the implications of our work for research and practice.

6.1. Design principles for explainable LLM-based open data assistants

In this section, we synthesize the empirical findings from our two-cycle DSR project into two generalizable design principles (DPs) for explainable LLM-based open data assistants. While additional principles

could guide the design of such assistants, we focus our DPs specifically on the two most critical barriers to effective use identified in our evaluations: the lack of transparency in AI reasoning and the information overload caused by traditional data catalogs. To formalize our contributions, we adopt the schema proposed by Gregor et al. [25] which consists of four components: (1) aim, implementer, and user; (2) context; (3) mechanism; and (4) rationale. We extend this schema by introducing a fifth element—social impact—in order to make explicit how the DPs contribute to addressing social problems and delivering broader social good. As the implementer, target user base, application context, and social impact remain consistent across both principles, we define these shared elements below.

Our DPs are aimed at providers of open data portals and their respective developers as **implementers**. This group typically includes decision-makers in the IT department of public institutions. Regarding **users**, our findings from Cycle 1 underscore the need to design for a heterogeneous audience. While the user base includes domain experts and civil servants, our DPs primarily focus on citizens. They are characterized by diverse backgrounds and often lack the technical proficiency and domain knowledge to effectively use open data portals to satisfy their information needs. A typical example involves citizens tracking how their tax dollars are spent to be able to hold institutions accountable. Finally, the **context** of our research is anchored in governmental open data initiatives aimed at fulfilling the UN’s SDGs of transparent and accountable institutions. Specifically, we address municipal and state-level open data portals. Within this context, implementing our DPs can provide **social impact** by empowering citizens to hold their government accountable, which in turn increases perceived governmental transparency and trust. We believe our DPs may be applicable beyond this context to situations where other types of data need to be made accessible to diverse audiences. For example, LLM-based data assistants could support non-IT professionals within organizations who are less tech-savvy but still seek to independently analyze data and make data-driven business decisions [49]. Collectively, this interplay between implementers, users, context, and social impact forms the foundation of our DPs. We outline the empirical grounding of these principles below and summarize their aim, mechanism, rationale, and societal impact in Table 2.

6.1.1. Principle of transparent reasoning

A central finding of our DSR project is that open data assistants need to support users in understanding the analysis process rather than just providing answers. In our first design cycle, we observed that while technical transparency (e.g., displaying SQL queries) offers a form of proof, it fails to support non-technical users in understanding the assistant’s reasoning. Our second cycle demonstrated that natural language explanations of the reasoning process increase users’ ability to achieve representational fidelity. In other words, by describing the steps taken to arrive at an answer, the open data assistant helps users bridge the gap between their question and the complexity of the data underlying the assistant’s answer. This does not merely serve as a validation function; rather, it provides users with the context necessary

Table 2
Our design principles for explainable LLM-based open data assistants.

	Aim	Mechanism	Rationale	Societal Impact
Principle of Transparent Reasoning	Enable citizens to verify the open data assistant's answers	Supplement the assistant's final answers with natural language explanations that explicitly detail the intermediate analysis steps	By exposing the "how" of the data analysis in everyday language, the assistant shifts from a "black box" to a verifiable tool	Making complex open government data easier to access and understand empowers citizens to hold governments accountable, thereby increasing transparency and building trust in public institutions
Principle of Interface Simplification	Reduce information overload and prevent visual clutter that distracts citizens	Establish the conversational user interface as the default interaction mode and hide traditional, complex data catalogs (raw tables) by default	By shielding citizens from the complexity of raw data, the assistant helps them focus on the insights relevant to their specific questions	

to assess whether the analysis actually addressed their information needs. As a result, users are empowered to independently verify facts and hold government institutions accountable.

6.1.2. Principle of interface simplification

Our second principle derives from an unexpected finding in our final evaluation: instead of increasing users' ability to achieve representational fidelity, using a conversational assistant in combination with a traditional data catalog actually reduced it compared to when the catalog was completely hidden. This challenges a prevailing assumption in open data initiatives: that more data equals more transparency. While open data portals usually provide comprehensive data catalogs to maximize transparency, literature acknowledges that citizens often lack the relevant technical and analytical skills to derive insights from raw data [5,6]. Our analysis suggests that as users interacted more with the data catalog tables, representational fidelity decreased, indicating that the complexity of data catalogs may lead to information overload. This finding highlights that rather than simply adding an AI assistant to an existing open data portal, the overall user interface should be simplified to present only the information directly relevant to citizens' questions. As a result, the assistant could serve as an abstraction layer that simplifies complex data, thereby increasing transparency.

6.2. Implications for research and practice

Our findings have several implications for research and practice. First, our work contributes to the growing body of literature on the design of human-AI systems [50] by providing a novel artifact and design knowledge for explainable LLM-based open data assistants. Specifically, we move beyond traditional XAI approaches that typically focus on output justification. Instead, we show that providing natural language explanations of the open data assistant's reasoning traces, such as the selected tables and generated SQL queries, increases users' ability to obtain faithful representations and ultimately improves their effectiveness in finding answers to their questions. This confirms that the inherent 'black box' nature of LLM-based systems constitutes a barrier to effective use. By exposing the internal reasoning process in a comprehensible format, we enable users to critically assess the system's output for clarity, completeness, and meaningfulness.

Second, we shed light on potential tensions in the design of open data portals, which currently often feature a data catalog view to maximize transparency by allowing users to browse and download raw data. However, in contrast to our expectations, the simplified conversational UI led to higher levels of representational fidelity than the hybrid interface (with conversational UI and data catalog), challenging the common assumption that providing more information leads to higher representational fidelity. This phenomenon, which we refer to as the "transparency paradox", suggests that transparency is often mistakenly

equated with the mere technical availability of data, rather than its accessibility to a non-technical audience. This unexpected outcome serves as a prime example of design indeterminacy, a recognized challenge in DSR, in which theoretical prescriptions may not yield the predicted outcomes due to the complexity of IS artifacts [51]. More broadly, this finding underscores the importance of designing for the intended audience in the context of open data. While experts may rely on the data catalog for verification, citizens appear to achieve a better understanding when the assistant acts as an abstraction layer, shielding them from the complexity of raw data structures.

Third, our research contributes to TEU, the theoretical foundation guiding our artifact design, by revealing that the benefits of increased representational fidelity go beyond effective use to include broader societal impacts. While Burton-Jones and Grange [3] focus on effectiveness and efficiency as the primary outcomes of representational fidelity, our findings demonstrate that, in the context of government open data, it also positively impacts perceived government transparency and increases trust in the government. By establishing this causal chain—from obtaining faithful representations to perceived transparency to trust—we provide evidence that successful UI design contributes to broader social goals, such as the UN's SDG 16 regarding transparent and accountable institutions.

Finally, our research provides actionable guidance for practitioners (e.g., providers and developers of open data portals) and can be generalized to the design of LLM-based data assistants that serve diverse user groups across different contexts. The derived design principles offer concrete recommendations for implementing LLM capabilities and explanations that support the effective use of open data portals. Namely, the Principle of Transparent Reasoning (DP1) guides implementers to provide natural language explanations of the assistant's analytical steps. More critically, the Principle of Interface Simplification (DP2), derived from our finding that the traditional data catalog can negatively impact representational fidelity by causing information overload, recommends that developers use the conversational assistant as the primary view and hide complex tables by default. Public institutions and open data providers can follow these principles to enhance the accessibility and impact of their portals, ultimately advancing the goals of open data initiatives. If time and resources are limited, we suggest a two-step approach: first reducing complexity by simplifying the UI (DP1), then adding an LLM-based data assistant that enables users to ask questions about the data (DP2). Our findings also have implications for policymakers and open data advocates. The results suggest that LLM-based data assistants could help address the persistent challenge of underutilized open datasets by making them more accessible to non-technical users. This aligns with the UN's SDG 16, particularly its target of ensuring public access to information.

6.3. Limitations and future work

Despite these contributions, our research also has some limitations that provide opportunities for future work. First, we conducted a summative evaluation of our LLM-based open data assistant as a standalone system in a controlled experiment. While we integrated the data catalog view to mimic open data portal functionality, it raises the question of how such an assistant could be integrated into existing portals and affect citizens' effective use under real-world conditions. Therefore, conducting a field experiment in which real users interact with an open data portal featuring an integrated assistant could further validate our results and improve generalizability.

Second, the design of our open data assistant comes with limitations. Our explanation design was relatively static in nature. While the content was dynamically generated based on user input, the assistant's analytical process, and the underlying data, it offered limited interaction possibilities (primarily allowing users to request more detailed information). Further, we acknowledge that while we shed light on the "transparency paradox", our design does not offer a solution to resolving it. Our current approach—hiding the data catalog—prioritizes simplicity but potentially compromises direct verifiability. We therefore call for future research to develop novel interaction patterns that provide more sophisticated explanation and exploration mechanisms that enable users to explore alternative analysis paths or customize explanation depth according to their technical expertise.

Third, we acknowledge the inherent risks associated with the probabilistic nature of generative AI, including potential hallucinations and bias. While our technical evaluation confirmed that the assistant provides accurate answers approximately 80% of the time, users in our experiment only achieved an average effectiveness of 57%. This performance gap highlights a critical area for future investigation: while natural language lowers access barriers, effective use remains highly dependent on users' ability to formulate precise prompts—a skill that many users seem to lack. Furthermore, the system's high baseline accuracy could suggest an alternative explanation for our findings: the detailed explanations may have simply reassured users, leading them to accept the answers without deep critical reflection rather than actually improving their ability to achieve representational fidelity. Future research could therefore expand on our findings by examining user behavior when the data assistant provides incorrect answers, determining whether explanations truly improve effectiveness or induce overreliance.

7. Conclusion

In an era of increasing misinformation, declining democratic engagement, and growing distrust in public institutions, providing citizens with easy access to government information is more important than ever. Existing open data portals designed for this purpose, however, regularly fall short of this goal and do not enable citizens to hold elected leaders accountable or participate meaningfully in the democratic process. Our research presents a solution to this societal problem: explainable LLM-based open data assistants. These assistants not only enable citizens to ask questions in their own words and receive data-driven insights without requiring any technical skills, but also explain how they arrive at their answers in plain language. As a result, finding the information they need becomes much easier for citizens, which increases transparency and ultimately improves trust in government. We believe that our DSR project suggests just one way in which theory-driven design artifacts can address the many challenges faced by democratic societies today, and we encourage our colleagues to get involved in any way they can to help protect and strengthen democracy.

CRedit authorship contribution statement

Till Carlo Schelhorn: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Ulrich Gnewuch:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Alexander Maedche:** Writing – review & editing, Validation, Supervision, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Gemini, ChatGPT and Grammarly in order to improve the grammar and clarity of the text. After using these tools, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

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

References

- [1] U. Nations, The 17 Goals - Sustainable Development Goals.
- [2] R.P. Lourenço, An analysis of open government portals: A perspective of transparency for accountability, *Gov. Inf. Q.* 32 (3) (2015) 323–332.
- [3] A. Burton-Jones, C. Grange, From use to effective use: A representation theory perspective, *Inf. Syst. Res.* 24 (3) (2013) 632–658.
- [4] A. Quarati, M. De Martino, Open government data usage: a brief overview, in: *Proceedings of the 23rd International Database Applications & Engineering Symposium*, 2019, pp. 1–8.
- [5] I. Safarov, A. Meijer, S. Grimmelikhuijsen, Utilization of open government data: A systematic literature review of types, conditions, effects and users, *Inf. Polity* 22 (2017) 1–24.
- [6] A. Purwanto, A. Zuiderwijk, M. Janssen, Citizen engagement with open government data: A systematic literature review of drivers and inhibitors, *Int. J. Electron. Gov. Res.* 16 (2020) 1–25.
- [7] M. Ruoff, U. Gnewuch, A. Maedche, B. Scheibehenne, Designing conversational dashboards for effective use in crisis response, *J. Assoc. Inf. Syst.* 24 (6) (2023) 1500–1526.
- [8] E. Sánchez-Nielsen, A. Morales, O. Mendo, F. Chávez-Gutiérrez, SuDaMa: Sustainable open government data management framework for long-term publishing and consumption, *IEEE Access* 9 (2021) 151841–151863.
- [9] C. Stein, T. Teubner, S. Morana, Designing a conversational agent for supporting data exploration in citizen science, *Electron. Mark.* 34 (1) (2024) 23.
- [10] M. Gomez-Vazquez, J. Cabot, R. Clarisó, Automatic generation of conversational interfaces for tabular data analysis, in: *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 2024, pp. 1–6.
- [11] K. Gu, R. Shang, T. Althoff, C. Wang, S.M. Drucker, How do analysts understand and verify AI-assisted data analyses? in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.
- [12] K. Gu, M. Grunde-McLaughlin, A. McNutt, J. Heer, T. Althoff, How do data analysts respond to AI assistance? A wizard-of-oz study, in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.
- [13] J. Guo, V. Mohanty, J.H. Piazentin Ono, H. Hao, L. Gou, L. Ren, Investigating interaction modes and user agency in human-LLM collaboration for domain-specific data analysis, in: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–9.
- [14] F. Leiser, S. Eckhardt, V. Leuthe, M. Knaeble, A. Mädche, G. Schwabe, A. Sunyaev, HILL: A hallucination identifier for large language models, in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–13.
- [15] A.-M. Sirbu, T.C. Schelhorn, U. Gnewuch, Explanation provision strategies in LLM-based data assistants: Impact on extraneous cognitive load, trust, and task performance, *ECIS 2025 Proc.* (2025-06-18).
- [16] A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research, *MIS Q.* 28 (1) (2004) 75–105.
- [17] W. Kuechler, V. Vaishnavi, On theory development in design science research: anatomy of a research project, *Eur. J. Inf. Syst.* 17 (2008) 489–504.
- [18] S. Gregor, I. Benbasat, Explanations from intelligent systems: Theoretical foundations and implications for practice, *MIS Q.* 23 (4) (1999) 497–530.
- [19] V.-H. Trieu, A. Burton-Jones, P. Green, S. Cockcroft, Applying and extending the theory of effective use in a business intelligence context, *MIS Q.* 46 (2022) 645–678.

- [20] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, Y. Cao, React: Synergizing reasoning and acting in language models, 2023.
- [21] OpenAI, Learning to reason with LLMs, 2024.
- [22] M. Tremblay, A. Hevner, D. Berndt, S. Chatterjee, The Use of Focus Groups in Design Science Research, vol. 22, 2010, pp. 121–143.
- [23] T. Miller, Explainable AI is dead, long live explainable AI! hypothesis-driven decision support using evaluative AI, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, pp. 333–342.
- [24] K. Bauer, M. von Zahn, O. Hinz, Expl(AI)ned: The impact of explainable artificial intelligence on users' information processing, *Inf. Syst. Res.* 34 (4) (2023) 1582–1602.
- [25] S. Gregor, L.C. Kruse, S. Seidel, The anatomy of a design principle, *J. Assoc. Inf. Syst.* 21 (6) (2020).
- [26] S. Sadiq, M. Indulska, Open data: Quality over quantity, *Int. J. Inf. Manage.* 37 (3) (2017) 150–154.
- [27] A. Zuiderwijk, M. Janssen, Y.K. Dwivedi, Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology, *Gov. Inf. Q.* 32 (4) (2015) 429–440.
- [28] S. Neumaier, V. Savenkov, S. Vakulenko, Talking open data, in: The Semantic Web: ESWC 2017 Satellite Events, 2017, pp. 132–136.
- [29] S. Porreca, F. Leotta, M. Mecella, S. Vassos, T. Catarci, Accessing government open data through chatbots, in: Current Trends in Web Engineering, 2018, pp. 156–165.
- [30] S. Keyner, V. Savenkov, S. Vakulenko, Open data chatbot, 2019.
- [31] R. del Hoyo-Alonso, V. Rodrialvarez-Chamarro, J. Veá-Murgía, I. Zubizarreta, J. Moyano-Collado, Aragón open data assistant, lesson learned of an intelligent assistant for open data access, in: Conversations, 2023.
- [32] E. Karpas, O. Abend, Y. Belinkov, B. Lenz, O. Lieber, N. Ratner, Y. Shoham, H. Bata, Y. Levine, K. Leyton-Brown, D. Muhlgay, N. Rozen, E. Schwartz, G. Shachaf, S. Shalev-Shwartz, A. Shashua, M. Tenenholz, Mrkl systems: modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning, 2022, <http://dx.doi.org/10.48550/arXiv.2205.00445>.
- [33] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q.V. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, *Adv. Neural Inf. Process. Syst.* 35 (2022) 24824–24837.
- [34] T. Wu, M. Terry, C.J. Cai, AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–22.
- [35] J.S. Dhaliwal, I. Benbasat, The use and effects of knowledge-based system explanations: Theoretical foundations and a framework for empirical evaluation, *Inf. Syst. Res.* 7 (3) (1996) 342–362.
- [36] M.A. Meza Martínez, M. Nadj, M. Langner, P. Toreini, A. Maedche, Does this explanation help? Designing local model-agnostic explanation representations and an experimental evaluation using eye-tracking technology, *ACM Trans. Interact. Intell. Syst.* 13 (4) (2023) 27:1–27:47.
- [37] J. Schoeffer, M. De-Arteaga, N. Kuehl, Explanations, fairness, and appropriate reliance in human-AI decision-making, 2024.
- [38] K.Z. Gajos, L. Mamykina, Do people engage cognitively with AI? Impact of AI assistance on incidental learning, in: Proceedings of the 27th International Conference on Intelligent User Interfaces, 2022, pp. 794–806.
- [39] Q.V. Liao, J.W. Vaughan, AI transparency in the age of LLMs: A human-centered research roadmap, *Harv. Data Sci. Rev. (Special Issue 5)* (2024).
- [40] A. Schmitt, Ensuring human agency: A design pathway to human-AI interaction, *ICIS 2024 Proc.* (2024).
- [41] K.R. Larsen, R. Lukyanenko, R.M. Mueller, V.C. Storey, J. Parsons, D. VanderMeer, D.S. Hovorka, Validity in design science, *MIS Q.* (2024) 1–31.
- [42] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, H. Jégou, THE FAISS LIBRARY, *IEEE Trans. Big Data* (2025) 1–17.
- [43] D.L. Chen, M. Schonger, C. Wickens, Otree—An open-source platform for laboratory, online, and field experiments, *J. Behav. Exp. Financ.* 9 (2016) 88–97.
- [44] M. Mahmood, V. Weerakkody, W. Chen, The role of information and communications technology in the transformation of government and citizen trust, *Int. Rev. Adm. Sci.* 86 (4) (2020) 708–728, Publisher: SAGE Publications Ltd.
- [45] F. Bélanger, L. Carter, Trust and risk in e-government adoption, *J. Strat. Inf. Syst.* 17 (2) (2008) 165–176.
- [46] X. Zhao, J.G. Lynch Jr., Q. Chen, Reconsidering baron and kenny: Myths and truths about mediation analysis, *J. Consum. Res.* 37 (2) (2010) 197–206.
- [47] G.W. Cheung, R.S. Lau, Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models, *Organ. Res. Methods* 11 (2) (2008) 296–325, Place: US Publisher: Sage Publications.
- [48] T. Jetzek, M. Avital, N. Bjørn-Andersen, Generating value from open government data, *ICIS 2013 Proc.* (2013).
- [49] M. Hvalshagen, R. Lukyanenko, B. Samuel, Empowering users with narratives: Examining the efficacy of narratives for understanding data-oriented conceptual models, *Inf. Syst. Res.* (2022-08-31) <http://dx.doi.org/10.1287/isre.2022.1141>.
- [50] V.C. Storey, A.R. Hevner, V.Y. Yoon, The design of human-artificial intelligence systems in decision sciences: A look back and directions forward, *Decis. Support Syst.* 182 (2024) 114230.
- [51] R. Lukyanenko, J. Parsons, Research perspectives: Design theory indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? *J. Assoc. Inf. Syst.* 21 (5) (2020-09-11) <http://dx.doi.org/10.17705/1jais.00639>.

Till Carlo Schelhorn is a Ph.D. student at the Karlsruhe Institute of Technology (KIT) in Germany. He is part of the human-centered systems lab (h-lab), and his research focuses on designing LLM-based assistance for data analysis and business intelligence.

Ulrich Gnewuch is a Professor of Information Systems at the University of Passau, Germany, where he holds the Chair of Explainable AI-based Business Information Systems. His research focuses on the design, use, and impact of AI in business and society. His work has been published in journals such as *Information Systems Research*, *Journal of the Association for Information Systems*, and *Business & Information Systems Engineering*.

Alexander Maedche is a professor of Information Systems at the Karlsruhe Institute of Technology (KIT), Germany. He heads the human-centered systems lab (h-lab) at the Institute for Information Systems (WIN). His work is positioned at the intersection of Information Systems and Human-Computer Interaction (HCI). Specifically, his research is focused on designing human-centered systems for better work and life. His work has been published in leading international information systems and computer science outlets such as the *Management Information Systems Quarterly (MISQ)*, *Information Systems Research (ISR)*, *Decision Support Systems (DSS)* and the *ACM Conferences on Human Factors in Computing (CHI)* and on *Computer-Supported Cooperative Work & Social Computing (CSCW)*.