Review

Chatbot-Based Natural Language Interfaces for Data Visualisation: A Scoping Review

Ecem Kavaz [1,*] , **Anna Puig** [1,2] **and Inmaculada Rodríguez** [1,3]

1   Departament de Matemàtiques i Informàtica, Universitat de Barcelona (UB), 08007 Barcelona, Spain; annapuig@ub.edu (A.P.); inmarodriguez@ub.edu (I.R.)
2   Institut of Complex Systems (UBICS), Universitat de Barcelona (UB), 08007 Barcelona, Spain
3   Institut de Matemàtica UB (IMUB), Universitat de Barcelona (UB), 08007 Barcelona, Spain
*   Correspondence: ekavazka27@alumnes.ub.edu

Abstract: Rapid growth in the generation of data from various sources has made data visualisation a valuable tool for analysing data. However, visual analysis can be a challenging task, not only due to intricate dashboards but also when dealing with complex and multidimensional data. In this context, advances in Natural Language Processing technologies have led to the development of Visualisation-oriented Natural Language Interfaces (V-NLIs). In this paper, we carry out a scoping review that analyses synergies between the fields of Data Visualisation and Natural Language Interaction. Specifically, we focus on chatbot-based V-NLI approaches and explore and discuss three research questions. The first two research questions focus on studying how chatbot-based V-NLIs contribute to interactions with the Data and Visual Spaces of the visualisation pipeline, while the third seeks to know how chatbot-based V-NLIs enhance users' interaction with visualisations. Our findings show that the works in the literature put a strong focus on exploring tabular data with basic visualisations, with visual mapping primarily reliant on fixed layouts. Moreover, V-NLIs provide users with restricted guidance strategies, and few of them support high-level and follow-up queries. We identify challenges and possible research opportunities for the V-NLI community such as supporting high-level queries with complex data, integrating V-NLIs with more advanced systems such as Augmented Reality (AR) or Virtual Reality (VR), particularly for advanced visualisations, expanding guidance strategies beyond current limitations, adopting intelligent visual mapping techniques, and incorporating more sophisticated interaction methods.

Keywords: data visualisation; natural language interface; chatbot; survey

1. Introduction

Nowadays, the large increase in data generated by a wide myriad of sources, such as social media, scientific simulations and IoT sensors, has highlighted the need to make data more understandable [1]. In this context, data visualisation is essential for discovering data insights and identifying patterns, trends and outliers. Indeed, visual representations can transform raw data into meaningful stories that are easier for people to process and comprehend [1]. However, creating the right visualisations to help users easily understand the data is a challenging task. These representations should provide analysts with the appropriate parameters, layouts and interactions to explore huge and complex datasets, especially in terms of their size, the number of attributes (i.e., multidimensional data) and the relationships between them (i.e., correlations, dependencies, hierarchical relationships, network configurations) [2].

In recent years, a wide range of complex and multidimensional data visualisations have been proposed in the scientific community [1], either for specific datasets [3] or as more general visualisation methods [4,5], such as Sankey diagrams [6], Sunburst maps [5], tree maps [7] and network graphs [8]. But not only academic research is interested in data

analysis to improve or examine their work; many companies also rely on data analytics to improve their businesses and, hence, enhance the services provided to their users [9]. Consequently, visualisation methods and tools are evolving rapidly and constantly to solve new challenges posed in the field to adapt to changes.

Although static visualisations, i.e., those that do not have any interactive elements, are useful in certain circumstances, such as analysing simple data, this is not the case with large multidimensional data containing complex relationships that require more user interactions to navigate through the data [10]. Indeed, the complexity of the data and its multidimensionality requires a wide range of interaction possibilities to filter specific data, to show projections in 2D and 3D, to examine connections between data items and cluster them, and to obtain statistics, among others [11]. This complexity leads to the design of intricate visualisation systems with steep learning curves [12]. Non-expert users, who are not used to working with visualisations or analysis tools, can have particular difficulties selecting the visualisation method that best fits their data. Fortunately, advances in sensors and Natural Language Processing (NLP) technologies have facilitated the use of natural interaction methods based on body gestures and conversations that allow for the creation of seamless and comfortable user experiences [13].

Focusing on Visualisation-oriented Natural Language Interfaces (V-NLIs), many academic research projects and popular companies, such as Tableau, IBM Watson and Microsoft, have introduced and integrated them into their visualisation tools. These tools are effective and easy to learn, since they allow users to interact with visualisations using natural language, without needing to transform their queries into tool-specific actions and therefore allowing them to focus on their analysis [12]. In this context, natural language is considered a complementary input modality to direct manipulation (WIMP—Windows Icons Menus Pointers). In fact, the results of various studies have confirmed that users were more comfortable and interested in using multiple input modalities, i.e., multimodality [14,15]. Another major benefit of including Natural Language in visualisations is its inclusiveness [16], as it can support blind and low-vision people when interacting with visualisations.

Recently, large generative models such as ChatGPT [17] and DALL-E [18] have given a great impulse to the NLP field and surely may be exploited by V-NLI soon. However, NLIs (Natural Language Interfaces) still face major challenges. For instance, users' expectations are usually very high since they want to communicate with the system in the same way as they interact with other human beings. The conversational system therefore has to deal with ambiguities that might even be interpreted differently by different people [19].

Moreover, most NLIs for visualisation started out relying on limited jargon (i.e., vocabulary based on specific data), simple visualisations (e.g., bar charts, line charts) and functions such as filtering and selection. For instance, Cox et al., who were pioneers in the field, proposed a basic system using *form-based* interaction, meaning that users typed their queries (analytical intents) into a text box in order to obtain the corresponding visualisation outputs [20]. As research has advanced, more sophisticated V-NLIs have been developed, such as those referred to as *chatbot-based*. Chatbots are intelligent conversational systems that not only provide visual outputs to users but also guide them, especially users with less experience in visual analytics, with additional aids such as textual feedback, recommendations, and complex multi-stepped queries [21].

Despite V-NLI being a relatively new field, several survey papers have already addressed this topic. Shen et al. [12] presented a broad review of NLIs for visualisation. They summarised various features of NLIs including query interpretation, human interaction and dialogue management to highlight existing gaps in the field. Moreover, they analysed a variety of NLIs for visualisation including simple (one turn interactions), conversational (systems that track the conversation with follow-up questions) and narrative storytelling (systems that show multiple visualisations side by side with annotations).

Other reviews of the literature have focused on specific aspects of V-NLIs. Srinivasan et al. [11] proposed three task-based categories: visualisation-related tasks, data-

related tasks and system-control-related tasks. Moreover, a recent systematic review [22] analysed NLIs both for databases and for data visualisations in terms of input and output. On the input side, they examined multimodality and different types of queries, such as open-ended or factual. On the output side, they considered those that give textual answers, generate new visualisations and interact with existing ones.

In summary, previous works have focused on reviewing Visualisation-oriented Natural Language Interfaces that were mainly conceived as *form-based* question-answering systems, where the users ask the system questions using UI (User Interface) widgets, and the system's answer takes the form of text, a filtered visualisation and/or a new visualisation. Nevertheless, recent advances in Natural Language Processing have facilitated a double enhancement of these systems, both in its inner workings (NLU—Natural Language Understanding and NLG—Natural Language Generation) and in its interface. The interface is now a chatbot (embodied or not) that engages in conversation with the users to facilitate their interaction with visualisations. To the best of our knowledge, there has been no attempt in the V-NLI literature to specifically examine the relationship between the fields of data visualisation and chatbots. Thus, this paper presents a scoping review that analyses synergies between both fields and also summarises knowledge gained in analysing research works that have proposed *chatbot-based* V-NLIs for data visualisation. Our contribution is as follows:

- We present a scoping review to study the synergies between both data visualisation and chatbots fields to analyse how the use of chatbots improved data visualisation and visual analysis.
- We propose an analysis framework based on the three spaces of the data visualisation pipeline, i.e., Data Space, Visual Space and Interaction Space as well as on a characterisation of chatbots using four dimensions called AINT (A—Anthropomorphic, I—Intelligence, N—Natural Language Processing, and T—inTeractivity).
- We extract insights and challenges that will be helpful for researchers to develop and improve V-NLIs.

2. Background

In this section, we explore the two topics of this review, data visualisation and chatbot-based V-NLIs. We present the main vocabulary relating to these topics, which we will use to analyse them.

2.1. Data Visualisation

A common data visualisation process consists of several steps [12,23]. Figure 1 details the data flow through these steps constructing the visual structures and how the end-user can interact with the data involved in each step (from right to left, see the arrows in the lower part of the figure), filtering regions (**View Transformation**), changing visual parameters (**Visual Mapping**), and making more complex requests on the data (**Data Transformation**). Starting from the three spaces in which the visualisation takes place—Data Space, Visual Space and Interaction Space—we present the most relevant characteristics that will serve as a basis for describing the works under study in this scoping review.
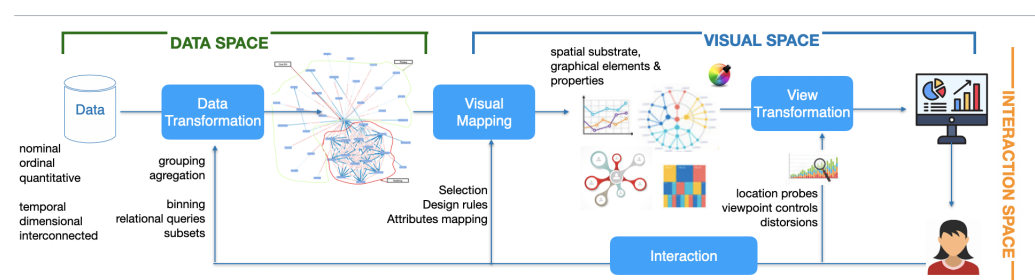

Figure 1. Overview of the Data Visualisation pipeline adapted from [12].

Data Space

The **Data Space** (shown in green in the upper-left part of Figure 1) covers the space in which the data are directly processed. When the input data are in a tabular format, the **Data Transformation** stage usually offers a set of operations to filter, cluster and aggregate data, among other functions, which can help to provide some data insights. To describe the related works in this systematic review, we use Shneiderman's [24] categories based on the implicit nature of the data, which are: data where items are distributed along the orthogonal axis (1D, 2D and 3D), data containing items in higher dimensionalities (complex use of the space when the dimension is greater than three), trees or hierarchical distributions (connected data), and networks (complex interconnected data). The two former categories are based solely on dimensionality, considering data as a set of individual items or sampled points in the space, in a structured or unstructured way, but without interconnections between them. However, trees and networks encode relationships between the sampled points: trees describe data containing parent–child relationships, while networks codify more complex relationships, which may be directed or undirected [25]. Moreover, in all the data categories, each point contains samples of different **attributes** that Shneiderman categorised as nominal, numerical (ordinal or quantitative) and temporal. Moreover, if these attributes are mapped into a 2D or 3D space, they are considered spatial.

Actually, these data categories help to identify the Data Transformation (see the first blue square in Figure 1), which is decisive for discovering insights in the data. Classical data transformations such as grouping, aggregation, enclosure and binning temporal items are widely associated with specific data categories in the visualisation community [26]. For instance, while aggregation functions such as mean and sum are suitable for quantitative data, grouping is better suited to nominal and ordinal data, and binning intervals is the right transformation in the case of temporal samples [12]. In addition, recent works have proposed more complex transformations of multidimensional datasets to extract meaningful subsets using relational queries [27–29]. In the case of connected structures, the topology can play an important role in the transformations, and also in the next stage of Visual Mapping [30]. For instance, extracting the largest path is a common transformation in elongated trees, and obtaining the widest level is a more typical transformation in compact hierarchies. Therefore, regarding the **data types** and their different transformations, in our study, we categorised data as: (1) **tabular data**, i.e., data with individual and non-connected items, where classical data transformations are enough, and (2) **complex data**, i.e., high-dimensional, temporal and interconnected data, which require more complex transformations. Moreover, both categories of data not only involve different transformations but also different strategies in the successive steps of the pipeline.

Visual Space

The second space involved in the data visualisation process is the **Visual Space** (shown in blue in the upper-middle part of Figure 1), which refers to how to map the data in visual structures (the Visual Mapping Step) and how to display them in a viewport (the View Transformation Step).

The Visual Mapping Step involves the definition of the next three aspects:

- The spatial substrate—i.e., the space and the layout used to map the data;
- The graphical elements—i.e., marks such as points, lines, images, glyphs, lines, etc.;
- The graphical properties—also called retinal properties, i.e., size, colour, orientation, etc. [23].

In the spatial substrate, a wide variety of layouts for displaying data have been proposed, from the simplest, such as those based on coordinate axes, to the more complex, such as those representing networks [31,32]. In fact, the more basic and simple they are, the more they are exploited in different applications. In our work, we classify these layouts as **basic** and **advanced**. Basic layouts refer to chart-based layouts, which have x and y axes (e.g., bar, line, scatter plot), table-based layouts and map-based layouts (such as a bubble map). We consider advanced layouts to be those that deal with higher dimensionalities

(e.g., parallel coordinates) and with connections (e.g., radial tree, circle packing, network graph, sunburst diagram, chord diagram).

Even with this simple classification into basic and advanced, we still have a wide range of basic and advanced layouts, and identifying the appropriate layout is therefore complex, especially if the users who analyse the data are not experts. Again, depending on the data types, some layouts fit better (i.e., a 3-aligned axis is a good choice to show quantitative spatial 3D data where each axis corresponds to one coordinate, and the circle packing layout fits well for simple hierarchical data). Moreover, once the layout has been selected, the next challenge is how to map the data attributes onto it. End-users can select and assign these characteristics manually, i.e., **user-defined** [33], but systems commonly use **pre-defined** layouts that only fit specific data. For example, Ref. [34] maps conversational hierarchical data with specific labelled attributes (e.g., negative or positive) to a stacked bar layout that is custom-designed for their data with indentations showing the hierarchy, and is therefore not flexible enough to be adapted to other data. Indeed, other approaches propose **rule-based** strategies to choose the layouts and their configuration dynamically according to the analysed data.

These rule-based approaches are commonly used in commercial systems such as PowerBI [35] and Tableau in [36]. Tableau integrated the "Show Me" algorithm [37], which selects and maps layouts depending on data type (text, date, date and time, numeric or boolean), data role (measure or dimension) and data interpretation (discrete or continuous). For example, to create a bar chart, users need to place at least one quantitative attribute and one categorical attribute to the y and x axes, respectively, and Tableau then automatically creates the bar chart. Similarly, Tableau needs two quantitative attributes to automatically create a scatter plot. Several academic studies used the "Show Me" algorithm to select visualisation methods [38–40]. Another rule-based method [30] deals with hierarchical data and infers the tree-based layout depending on the shape of the data hierarchy, i.e., they use tree layouts for elongated trees and radial layouts for compact structures. More **intelligent** approaches infer the most suitable layout using some visual examples given by users [41], while others recommend layouts from among five key design choices [42] and use pre-trained NN models that map data to predefined chart templates [28].

Additionally, the Visual Mapping step must consider which **graphical elements** to use and their properties. There is a broad range of graphical elements (also called mark types) used to map attributes, such as points, lines, glyphs, icons and symbols. Some of them are more suitable for displaying quantitative attributes such as points, while others are better suited to nominal data, where a symbol can communicate the meaning of the data in a pictorial way [2,43]. In this paper, we analyse the related works in terms of a semantic continuum of the graphical elements which goes from the more **abstract** (e.g., points, cross, stars) to the more meaningful or **symbolic** (e.g., glyphs, icons). We also take into account the graphical properties that can enhance one's understanding of the graphical elements, such as colours, size, position, orientation, value, textures, shapes, connectivity, grouping and animation. In addition, as in layout selection, finding adequate graphical elements for a given dataset and its properties is not a trivial task. In general, users can interactively select these graphical elements, although, as in the case of layouts, other methods have been proposed based on expert-defined rules [44] and intelligent algorithms that recommend [45] or infer the elements by means of pre-trained models with the most commonly used graphical elements [46]. Thus, in summary, to analyse the reviewed papers in terms of the **visual mapping identification**, i.e., to choose layouts and graphical elements and properties, we use the following categories: **fixed**, **user-defined**, **rule-based** (we refer to basic rule-based methods as those that follow a set of heuristics and make decisions based on them), and **intelligent methods** (intelligent methods involve the use of machine learning, artificial intelligence or other computational techniques to enable systems to learn from data, adapt and make decisions in a more flexible and adaptive manner).

Once the visual mapping is performed, the **View Transformation** stage allows users to change the viewpoint (e.g., zooming and panning), perform location probes (to measure

values in samples), and create some distortions in the image (i.e., change the projection type) [23]. Additionally, view transformation allows users to take into account multiple views simultaneously, as well as animations and others. Some view transformations emphasise data with importance-driven strategies to enhance values and regions of interest, among other factors. Focus+Context [47] highlights the important data (focus) while the rest of the data provide additional information on the background (context), which allows users to see the details as well as the entire perspective. For example, imagine a line chart showing sales over time in which the peak point is highlighted (focus) but you can still see the all sales over time in the background (context). Other methods use the size of the items to show different levels of detail simultaneously, such as the multi-resolution approach [48], which allows users to select different resolutions to drill down and see details as needed. For example, a treemap exploits multiresolution, showing overall sales of all the continents in the outer rectangles so that the user can select a specific continent to view the details of sales of its countries in nested rectangles. We will describe the reviewed works in terms of the number of views that they use simultaneously **(Single/Multiple)** and the **strategy** used to emphasise regions or parts of the view (zoom, panning, focus+context, level of detail, multiresolution and others).

Interaction Space

Last but not least is the **Interaction Space** (shown in blue in the upper-right part of Figure 1), where the users interact with all the previous steps defined above. There have been many attempts to categorise different interactions [49,50]. Yi et al. [10] proposed seven interaction methods based on the user's intents: **select**, **explore**, **reconfigure**, **encode**, **abstract/elaborate**, **filter** and **connect**. Select is used for marking data points choosing data points, layouts, while explore refers to navigating through the data, including functions such as zooming and panning. Reconfigure can be used to swap layout attributes on the x and y axis, or can use an algorithm to cluster some data points together in a network visualisation. Encode is used to assign or change graphical properties in terms of colour, size and shape. Abstract/elaborate displays details on demand such as collapsing/drilling down on a visualisation. The filter method shows data that fulfill a given condition. Finally, the connect method highlights the relationships between data items.

Users can utilise all these methods through different interaction styles. In our study, we consider a coarse two-labelled categorisation: Basic and Advanced. **Basic** styles refer to the WIMP (Windows, Icons, Mice, Pointer), while **Advanced** styles involve techniques such as Virtual Reality (VR), Augmented Reality (AR) and Natural Language. These categories will help us to explore the value that a visualisation-oriented chatbot can add to these interaction styles.

2.2. Chatbot

Chatbots are software systems able to engage in conversations with users [51], thereby representing a natural interface for them. This naturalness has favoured its spread in domains such as education [52], health [53], business [54] and, of course, fields such as visualisation analysis [55,56].

In Figure 2, we propose a general characterisation of chatbots using four dimensions, named AINT, depending on how we view them. First, chatbots may have Anthropomorphic (A) properties such as appearance [57] and gender and also may be endowed with personality and emotions [58]. Second, as an Intelligent system (I), task-based chatbots can proactively make data-driven decisions to give support to users' activities, and social chatbots maintain meaningful and engaging conversations with their users. In any case, chatbots can also be enhanced through a variety of AI methods and techniques, for example predicting users' necessities and behaviours and thereby personalising the UX (User eXperience) [59]. Third, as a Natural language processing system (N), chatbots usually consist of an NLU (Natural Language Understanding) part [60], which understands the intentions (goals) of the users (i.e., the inputs), maintaining the visual context of the conversation, but they must also provide a textual, visual, auditive answer to them, based on that context.

Those answer types (i.e., the outputs) can be either predefined or automatically generated. In the specific case of text, they are usually created by an NLG (Natural Language Generation) system [61]. Finally, as an interactive system (T), chatbots can be integrated with different interaction styles (WIMP, VR, XR) and be equipped with a multimodal interface through voice, text and gestures.
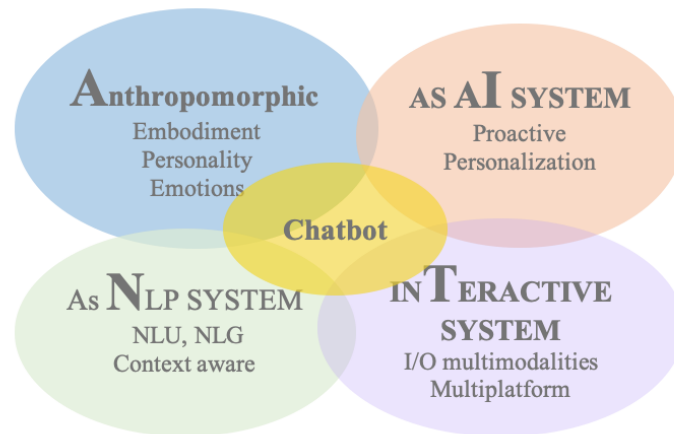

Figure 2. AINT—General characterization of a Chatbot based on four dimensions: A—Anthropomorphic, I—Intelligence, N—Natural Language Processing, and T—inTeractivity.

Next, we put our focus on chatbots in the specific context of visualisation. We analyse several aspects of the interactive space of a visualisation-oriented chatbot (see Figure 3), including its user interface as well as its input and output mechanisms, which are listed next to them in the figure, and will be explained in the following. From this analysis, there will emerge the main V-NLI features (User Interface, Input and Output, indicated in bold) and sub-features (indicated in italic) that will lead the analysis of the related work in this scoping review.
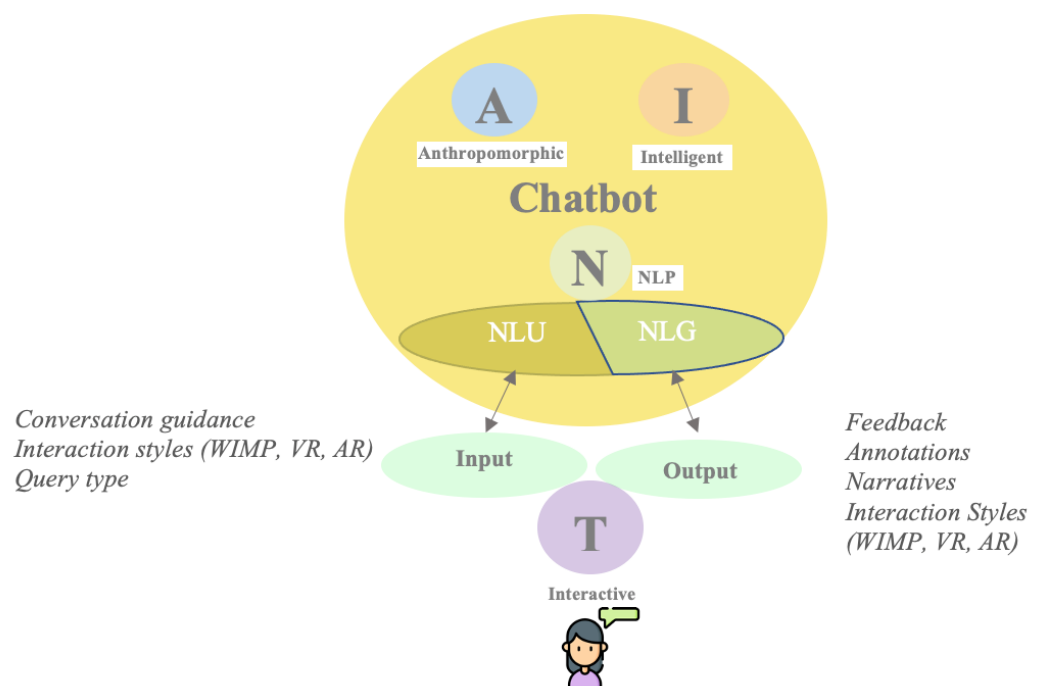

Figure 3. The interactive space's components of a V-NLI: User Interface, Input and Output.

User Interface

Visualisation-oriented Natural Language Interfaces (V-NLIs) are interactive systems (AIN**T**) designed to facilitate the users' visual analytic tasks. They can be designed using two different user interfaces (UI): a form-based interface and a chatbot-based interface. On one hand, a **form-based** V-NLI [40,62] (see Figure 4) is usually composed of a text box that allows the users to introduce the visualisation query using natural language, though it also has other widgets, for example, to refine (filter) the resultant visualisation. Nevertheless, these forms are usually not designed to engage in follow-up questions with the visualisation system. On the other hand, a **chatbot-based** interface [63] (see Figure 5) is distinguished by a named entity (also known as an agent), with gender and appearance, as well as with the ability to recognise and express emotions, while having personality traits (i.e., empathetic, fun, neutral). Chatbots are usually presented to the users as a separate "chat window" from the visualisations. This window displays the conversation but also complementary outputs (explanation, charts, and others), as we will see later. We can say that a chatbot-based V-NLI may have all of the aforementioned chatbot characteristics, i.e., AINT, meanwhile form-based V-NLI are potentially endowed with all of them except the anthropomorphic traits, i.e., INT.

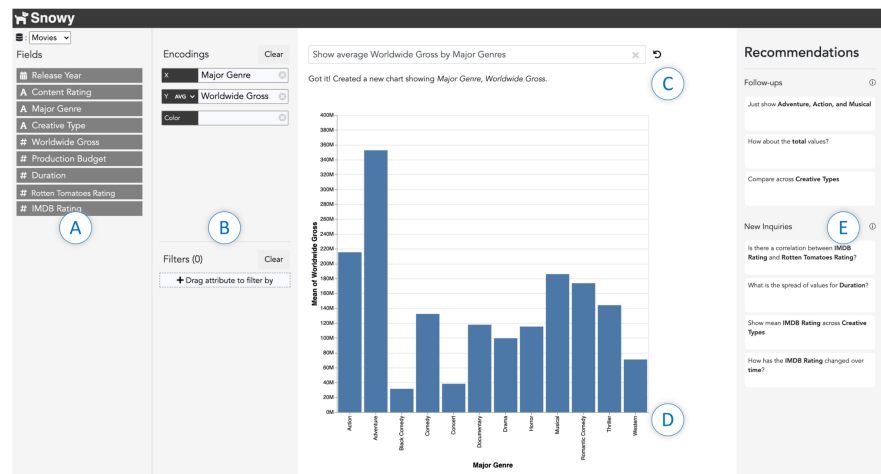**Figure 4.** Snowy [40], a form-based V-NLI example. Dashboard including: (A) Attribute panel, (B) manual view specification and filter panel, (C) NL input box and textual feedback, (D) visualisation space, and (E) query recommendation panel.
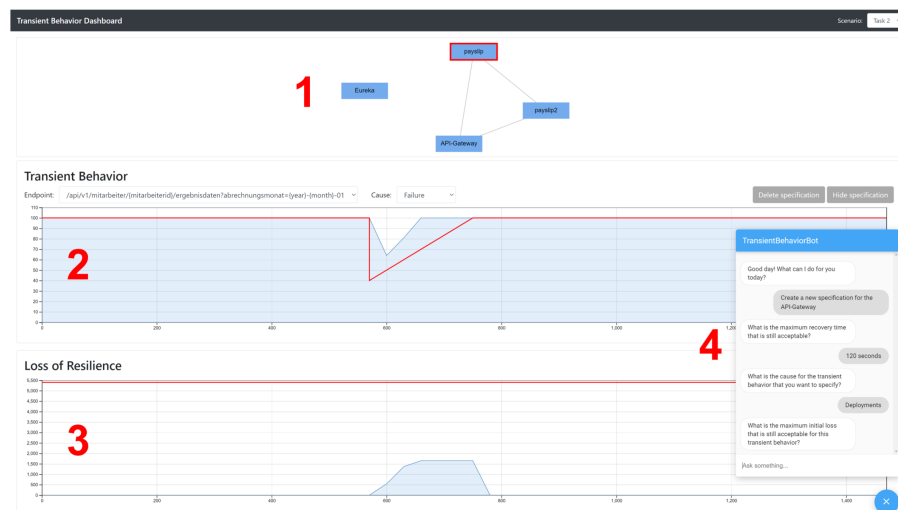

Figure 5. TransVis [63], a chatbot-based V-NLI example. Dashboard including; (1) Architecture visualisation, (2) and (3) area line graphs, and (4) chatbot window.

Input

The types of inputs (analytical questions) that a V-NLI system deals with are low- and high-level queries. In **Low-level** queries, the users explicitly describe their intent, for example, *"Show me action films that won an award in the past 10 years"*. Therefore, these queries can be interpreted easily. In contrast, **High-level** open-ended queries are naturally broader and their interpretation can be more complex [15,50]. In many cases, these high-level analytical questions should be decomposed as a series of low-level queries and be answered as such [64]. For example, to answer *"What are the trends in award-winning films?"* the system needs to infer the low-level queries: first, visualise award-winning films over a certain period of time, and then show their relevant characteristics (genre, special effects, franchises and others). Whenever the V-NLI system is not able to give an answer to this type of complex question, it might need to ask additional questions to the users. Note that both types of queries allow the users to interact with the data by means of the seven-interaction methods (select, explore, reconfigure, encode, abstract/elaborate, filter and connect) as defined in the description of the Interaction Space in Section 2.1, at any of the steps (View Transformation, Visual Mapping and Data Transformation) of the data visualisation pipeline as depicted in Figure 1.

Moreover, queries can be One-turn or Follow-up. In **One-turn** queries, the users ask the system in a single shot. Thus, even when the conversation may flow along several one-turn queries, it may not be necessary for the V-NLI system to maintain the context of the conversation [65]. On the other hand, the users usually perform **Follow-up** queries, which are a series of interconnected questions [66]. Therefore, the system should be able to remember the context of the conversation while answering the questions [15,39]. For example, if the first query is *"Colour nodes by age"* and the second query is *"Now by gender"*, NLI understands that the user continues talking about nodes and wants to use the same function, colour, but now colouring them by gender.

Nevertheless, the underlying system (AI**N**T) may fail to understand the queries of the users', thus not meeting their expectations [67]. Moreover, inexperienced users may have difficulties expressing their queries about visualisations. Then, the design of the conversational system faces the challenge of understandability, i.e., the ability of the system to be aware of the users' intents, really knowing and grasping the nuances of users' intentions, and also the challenge of discoverability [19], i.e., the ability of the users to know what they can ask to the system. Indeed, both properties are closely related since designing chatbots for discoverability may improve understandability.

The challenges of both understandability and discoverability require an interactive conversational system to guide the users on how to effectively communicate their goals (also referred to as intentions). Well-known **Conversational Guidance** strategies are based on **help**—the chatbot gives the users hints on what to ask; **intent auto-complete** functions—the system makes suggestions of possible intents while the users are writing the intent [62,68–70]; and **intent recommendations** [40]—after giving a response, the system suggests, based on data or on the previous turns of the analytical conversation, possible next intents to the users. Additionally, the understandability problem of NLIs is mainly derived from the biggest challenge that NL poses, which is ambiguity. One solution is to ask the users what they meant or to use disambiguity widgets [62,68]. For instance, when the user query is *"Show me medals for hockey"*, the NLI might not correctly interpret which type of hockey the user is referring to. Then, a widget may appear for the term *'Hockey'* showing two options *'indoor hockey'* and *'ice-hockey'* as both of these sports are basically called hockey. Thus, users can select the right one either through direct manipulation or by using natural language.

In the context of follow-up queries, the V-NLI should help the users to transition through the different visualisation states of the analysis. Indeed, research studies concluded that users prefer to carry out *analytical conversations*, meaning users want to go beyond the first visualisation they receive when making a request to the conversational interface [71]. Nevertheless, previous Conversational Guidance strategies (help, auto-complete and rec-

ommendations) may be sufficient for helping with the (initial) users' intent but could be insufficient for inferring the user's transitional intents (elaborate, adjust/pivot, start new, retry and undo) throughout the different visualisation states (interaction methods such as select attributes, filter, encode, transform) of an analytical conversation [67]. Therefore, *intelligent* Conversational Guidance (AINT) approaches are needed to predict users' goals based on their interactions throughout the analytical conversation and then proactively guide the user.

Another aspect of analytical visualisations is that they require dealing with co-reference, since the users may refer differently to the same entity during the conversation, for example, using pronouns. Fortunately, nowadays the most common NLP toolkits such as spaCy and AllenNLP incorporate components for co-reference resolution in their pipelines [72]. Moreover, an interesting case of co-reference arises when natural language interfaces coexist with other interaction styles (**Multimodality**) such as menu selection (WIMP—Window Icon Mouse Pointer) and direct manipulation (XR—Virtual or Augmented Reality) [73]. It may happen that users' NL queries refer to what they directly manipulated using clicks, gestures or eye-gaze, and, in consequence, the V-NLI system should keep also track of these non-textual references, i.e., the users' reference to what they did, not only to what they said. Thus, there should be a way of translating (WIMP, VR, AR) manipulations in the visualisation to text (with named entities) and so to be ready to be solved by the co-reference model.

Output

In addition to the requested visualisation, a V-NLI can consider **Complementary Output** such as **Feedback**, either **text** or **visual**: (i) to inform about the query's success or failure, (ii) to justify relevant decisions taken by the system, (iii) to provide the users with additional explanations to better interpret the resulting visualisation (textual, oral, graphs, or statistics) and annotations, and (iv) to display changes in the User Interface (highlighting menus, buttons). Specifically, annotations are superposed visual elements that enhance the generated visualisation and thereby further communicate more information [74]. Another type of complementary output is a visual narrative, which is text combined with images presenting the information with narrative components (actor, plot, setting) [75]. Finally, when there are other **Interaction Styles** integrated into the V-NLI system, the output should be synchronised to help the users be aware of the operation performed (e.g., filters updated in WIMP), and it should also be enhanced to facilitate a better understanding of the required visualisation (e.g., overlying images in AR) and to better communicate the system's response (e.g., haptic feedback in VR).

3. Method

We conducted this scoping review (a method that is used to analyse existing literature rapidly by mapping information using defined key concepts to find evidence and identify research gaps [76]) following the guidance article by Peters et al. [77] and we used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist, introduced by Tricco et al. [78]. That is, we first introduce our main objective, stating three research questions. Next, we explain the inclusion and exclusion criteria used to find the relevant works in the area, as well as the searching strategy. Finally, we describe the categories we have selected to analyse the compiled studies.

Note that we considered different PRISMA recommendations as follows. First, regarding the publication bias, we conducted a comprehensive and non-selective search across multiple databases, including searching for unpublished studies and personal communication with researchers to obtain complete information about relevant studies. Second, related to language bias, although PRISMA recommends that systematic reviews should not be limited to studies published in a specific language, a limitation of our study is the selection of English-only papers because of both the limited resources for translation and the lack of a comprehensive non-English literature. Third, for a future updating of this review, we

propose to revise the search strategy, reassess inclusion and exclusion criteria, conduct a new quality assessment, and update the data analysis. Finally, notice that we do not have any conflict of interest that could potentially bias the objectivity or impartiality of our review.

3.1. Objectives

The main objective of this scoping review is to systematically map the research conducted in chatbot-based NLIs for data visualisation. We primarily focus on answering the question, *"How has the use of chatbots improved data visualisation and visual analysis?"* In this context, we want to analyse specifically the synergies between both fields of data visualisation and chatbots based on the three spaces of the data visualisation pipeline, i.e., Data Space, Visual Space and Interaction Space (see Figure 1). Therefore, we will describe and summarise the scientific evidence on this topic to identify any existing gaps and offer future research directions. To further clarify our goal, we explore three research questions related to each space.

- RQ1: How do chatbot-based V-NLIs contribute to interactions with the Data Space?
- RQ2: How do chatbot-based V-NLIs contribute to interactions with the Visual Space?
- RQ3: How do chatbot-based V-NLIs enhance the user's interaction with the visualisation?

3.2. Study Selection

In this review, articles were selected according to the inclusion criteria, which we define as the following: (i) Articles that are related to Natural Language Interfaces with data visualisations; (ii) Articles that are written in English; (iii) Articles that are published between 2000 and 2023; and (iv) Articles long or short that are published in a journal, conference or book chapter. Furthermore, articles were excluded if (i) Articles included NLI or visualisation individually, i.e., NLI used to answer questions directly from a database or visualisations lacking NL input modality; (ii) Full text of the article was not available; and (iii) Articles that did not present and contribute original work (i.e., opinion articles).

3.3. Sources of Evidence and Search Strategy

The search strategy was developed according to the three-step JBI [79] standard approach recommended for scoping reviews:

- **Step 1.** Limited search to refine initial keywords: to find related articles, we searched databases (IEEE Xplore, the ACM Digital Library and Springer) with a combination of keywords including {('chatbot') AND ('visualisation')}, {('natural language interface') AND ('visualisation')}. We found a total of 3550 records in this step of the search (IEEE Xplore: 473, ACM: 525, Springer: 2552).
- **Step 2.** Search with refined keywords on Google Scholar, in relevant conferences (EuroVis, IEEE VIS, CHI) and research groups of the area, as well as recent surveys and systematic reviews: ('chatbot') AND ('data visualisation')}, {('natural language interface') AND ('data visualisation')}, {('conversational agent') AND ('data visualisation')}. We found 3148 references without duplicates.
- **Step 3.** Hand-refined search of found references: we screened titles and abstracts of the papers selected in the first two steps, and, if necessary, we reviewed the full text. In the final selection, excluding surveys, reviews and poster papers, we identified 62 recent articles from selected sources that are about the Natural Language Interface for Data Visualisation (V-NLI). However, as we focus concretely on chatbot-based V-NLIs, we excluded 42 of these articles, and we selected a total of 20 related articles.

3.4. Data Extraction

To analyse the collected works, we use the categories defined in Section 2, exploring the three spaces involved in the data visualisation pipeline (Data, Visual and Interaction Spaces), as well as chatbot characteristics (the interface, the Input and the complementary Output). In the following, we summarise the categories detailing the tables where the reviewed works are analysed.

Categories related to the Data Space (see Table 1) include:

- Description of data.
- Data type (Tabular or Complex).
- Attributes (Nominal, Numerical, Temporal, Spatial).
- Data transformation.

In relation to the Visual Space (see Table 2), the characteristics considered include:

- Visualisation Category (Basic or Advanced) and Type.
- Abstract (Lines, Points, Bar) and Symbolic (glyphs, icons) graphical elements.
- Visual Mapping Identification (Fixed, User-defined, Rule-based, Intelligent).
- View Transformation (Single or Multiple).
- Interaction Style (Basic—WIMP or Advanced—NL) .

In the Interaction Space, we collect information about the seven interaction methods proposed by Yi et al. [10]: select, explore, reconfigure, encode, abstract/elaborate, filter and connect.

Finally, regarding chatbots:

- V-NLI Interface (chatbot-based or form-based).
- Input:
  – Query Type (low or high);
  – One-turn or Follow-up queries;
  – Conversational Guidance (Help, Auto-complete, Recommendation);
  – Multimodality (WIMP, Touch, Gestures).
- Output:
  – Feedback (textual or visual): inform, justify, decisions, additional explanations (text, oral, graph, statistics, annotations), UI changes (menus, buttons) and narratives.
  – Interaction Style (WIMP, VR, AR).

Table 1. Summary of V-NLIs (ordered alphabetically by name) in defined data categories. Description of data, Data type (tabular and complex), Attributes (Nominal (Nom), Numerical (Num), Temporal (Temp), and Spatial (Spat)), and Data Transformation.

V-NLI	Year	Data Space			
		Description of Data	Data Type	Attributes	Data Trans.
ACUI [80]	2017	Software bundles	Complex (Network)	Nom	
Ava [81]	2020	Data science	Tabular	Num	✓
Boomerang [82]	2021	Finance	Tabular	Nom, Num	✓
Chat2Vis [83]	2023	Movies, Cars, etc.	Tabular	Nom, Num	
ConVisQA [69]	2020	Conversations	Complex (Hierarchical)	Conversation	
Data@Hand [84]	2021	Health metrics	Complex (Temporal)	Num, Temp	✓
DataBreeze [85]	2020	Colleges	Tabular	Nom, Num	
Evizeon [39]	2017	Diseases, houses	Tabular	Nom, Num, Spat, Temp	✓
FlowNL [70]	2022	Hurricanes	Complex (Flow)	Hurricanes, Spat, Num	
GameBot [86]	2020	Sports data	Tabular	Nom, Num, Temp	
GeCoAgent [87]	2021	Diseases	Tabular	Nom	✓
InChorus [88]	2020	Finance	Tabular	Nom, Num, Temp	✓
Iris [89]	2018	Data science	Tabular	Nom, Num	✓
MIVA [90]	2020	Coronavirus data	Tabular	Nom, Num, Spat, Temp	
ONYX [91]	2022	Coronavirus data	Tabular	Nom, Num, Spat, Temp	
Orko [15]	2017	Football players	Complex (Network)	Nom, Num	
Snowy [40]	2021	Movies	Tabular	Nom, Num, Temp	✓
Talk2Data [64]	2021	Finance, Cars	Tabular	Nom, Num, Temp	✓
TransVis [63]	2021	Transient	Complex (Network)	Nom, Num	
Valetto [92]	2018	Cars	Tabular	Nom, Num	✓

Table 2. Summary of V-NLIs in defined visualisation categories. Visualisation Category (Basic and Advanced), Graphical Elements (Lines, Points, Bars), Visual Mapping Identification (Fixed, User-defined, Rule-based and Intelligent), and View Transformation (Single and Multiple views).

V-NLI	Visual Space			View Trans.
	Visual Mapping			
	Visualisation Category (Type)	Graphical Elements	Visual Mapping Identification	
ACUI [80]	Adv (Network)	Lines, Points	Fixed	Single
Ava [81]	Basic (Line)	Lines	User-defined	Single
Boomerang [82]	Basic (Bar, Scatter, Line)	Lines, Points	Rule-based	Multiple
Chat2Vis [83]	Basic (Bar, Scatter, Line, Box-plot)	Lines, Points, Bars	Intelligent & User-defined	Multiple
ConVisQA [69]	Adv (Hierarchical stacked bar)	Bars	Fixed	Single
Data@Hand [84]	Basic (Bar, Line)	Lines, Bars	Fixed	Multiple
DataBreeze [85]	Basic (Dots)	Points	Fixed	Single
Evizeon [39]	Basic (Bar, Scatter, Line, Map)	Lines, Points, Bars	Fixed & Rule-based	Multiple
FlowNL [70]	Adv (Flow) & Basic (Bar, Map)	Lines, Bars	Fixed	Multiple
GameBot [86]	Basic (Bar, Line, Table, Shot)	Lines, Points, Bars	Rule-based	Single
GeCoAgent [87]	Basic (Pie)	Pies	Fixed	Single
InChorus [88]	Adv (Parallel) & Basic (Bar, Scatter, Line)	Lines, Points, Bars	Rule-based & User-defined	Single
Iris [89]	Basic (Scatter)	Points	User-defined	Single
MIVA [90]	Basic (Bar, Line, Map)	Lines, Points, Bars	Fixed	Multiple
ONYX [91]	Basic (Bar, Scatter, Map)	Points, Bars	Fixed & User-defined	Single
Orko [15]	Adv (Network) & Basic (Bar)	Lines, Points, Bars	Fixed	Multiple
Snowy [40]	Basic (Bar, Scatter, Line)	Lines, Points, Bars	Rule-based	Single
Talk2Data [64]	Basic (Bar, Scatter, Line, Pie, Area)	Lines, Points, Bars, Pies	Rule-based	Multiple
TransVis [63]	Adv (Network) & Basic (Line)	Lines, Points	Fixed	Multiple
Valetto [92]	Basic (Scatter)	Points	Fixed	Single

4. Results

In the 20 articles identified by our search strategy (Section 3.3), we included 10 articles that are chatbot-based [63,80–83,86,87,89,91,92] and 10 articles that are form-based NLIs but have some chatbot characteristics such as providing feedback [15,39,40,62,64,69,70,84,88,90].

4.1. Data Space

We analysed all the research works in terms of the main characteristics involved in the **Data Space**; see Figure 6.
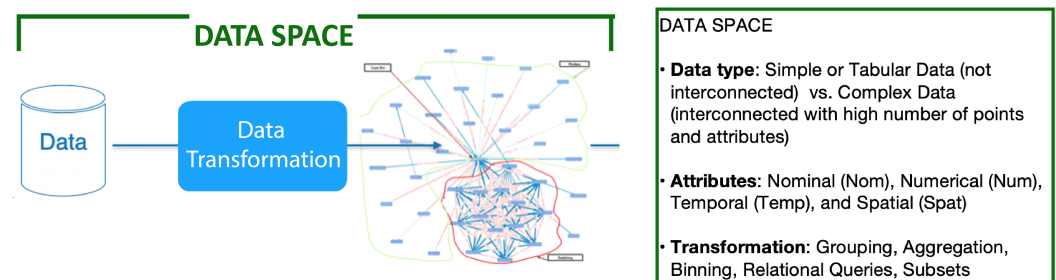

Figure 6. Data Space overview and the main characteristics of the data involved in the visualisation pipeline.

Table 1 summarises the analysed research works describing the explored data through visualisation: movies, sports, coronavirus, finance and others. We found that 70% (14/20) of them used multidimensional *tabular* data [39,40,64,81–83,85–92], while some of them also included spatial data [39,70,90,91]. Moreover, the table details the kind of data attributes each V-NLI supports (nominal, numerical, temporal, spatial). Six of the twenty visualisation systems used *complex data*. For instance, Ref. [80] has data related to software bundles and

services such as OSGi bundles, and [15] uses network data displaying the relationships between football players. Furthermore, Ref. [69] has hierarchical data that is collected from online conversations and [70] works with flow data such as hurricanes. Finally, Ref. [84] has sequential temporal data (e.g., sleep time during each night), and [63] has transient data which is a data type that is relevant to a time period; in this case, it is the quality of software services over time.

Regarding the **Data Transformation** step, 10/20 V-NLIs used a kind of data transformation. For example, Ava [81] and Iris [89] are both designed to facilitate data science tasks and they transform data to perform statistical analyses, such as logistic regression and finding correlations, respectively. Similarly, Valetto [92] and Boomerang [82] compute correlation between attributes, and the latter also finds aggregated values. Data@hand [84], InChorus [88], Evizeon [39] and Snowy [40] also perform aggregation functions such as average and sum. GeCoAgent [87] also computes aggregation functions, as well as other data transformations such as clustering, regression, etc., while extracting genomics data. Finally, Talk2Data [64] calculates the difference between numerical attributes. In general, the reviewed methods applied basic transformations that were not highly complex, i.e., those that entail more "intelligent" data processing.

4.2. Visual Space

Regarding the **Visual Space** (Figure 7), we summarise in Table 2 the reviewed works in terms of Visual Mapping components (graphical elements, identification and type of visualisation or substrate) and View Transformation details (use of single or multiple views and the type of interactions). Most of the articles employed explicitly *basic* visualisations (14/20, 70%) [39,40,64,81–87,89–92]. The most common methods are bar charts, line charts and scatter plots.
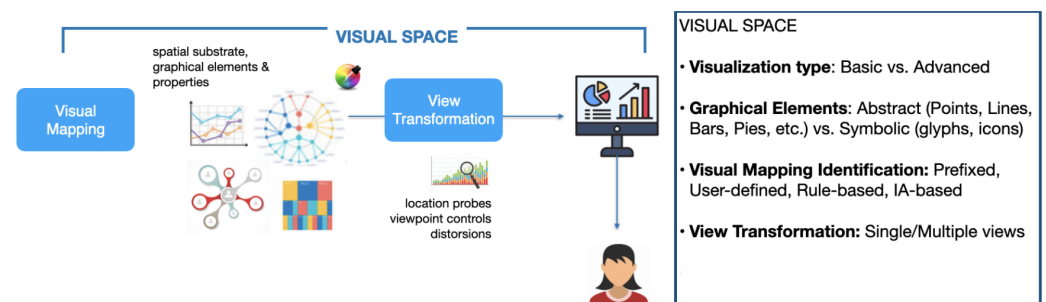

Figure 7. View Space overview and the main characteristics of the Visual Mapping and the View Transformation steps.

Most of the V-NLIs include a combination of these common methods [40,64,82–84,86]. Additionally, Chat2Vis [83] includes a box plot, Talk2Data [64] has a pie chart and Gamebot [86] offers users game shot charts about football and basketball games, as well as displays games statistics in tables. Moreover, some V-NLIs include a map chart, while [39,90,91] have 2D maps in addition to the most popular methods, and [70] includes a 3D map. There are five V-NLIs that have only one visualisation method: GeCoAgent [87] has a pie chart, Ava [81] has a line chart, Valetto [92] and Iris [89] have scatter plots, and DataBreeze [85] uses dots to visualise every data point individually.

Only a small percentage of these studies used *advanced* visualisation methods (6/20, 30%) (see Figure 8). For example, Bieliauskas and Schreiber [80] and Orko [15] implemented network visualisations as their main visualisation. In addition, Orko includes additional basic visualisation methods such as a bar chart to support its main visualisation. On the other hand, Tansvis [63] uses a line graph as the main visualisation for analysing transient data (quality of the software system over time), though it has a network graph for displaying the overview of the software system, where users can select a part to explore transient behaviours. ConVisQA [69] created a novel design to show the hierarchical structure of conversations using stacked bar charts with indentations to show the hierarchy. ConVisQA

also displays the conversations on the right-hand side of the screen. InChorus [88] supports popular basic visualisations such as bar, line and scatter, and they also included one complex option, parallel plots. FlowNL [70] used flow visualisation to show flows occurring on the earth (e.g., hurricanes). FlowNL also included basic visualisation methods for giving additional information such as a bar chart displaying the velocity of the hurricanes.
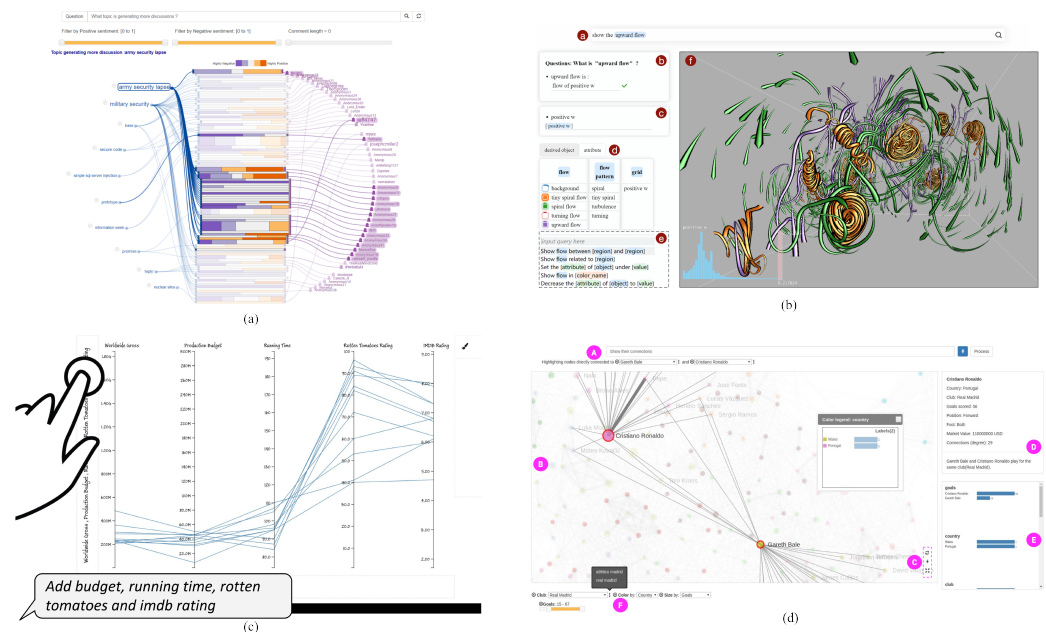

Figure 8. Four proposals for complex visualisation: (**a**) ConVisQA [69], (**b**) FlowNL [70] including (a) input box, (b) dialog box to solve unknown terms, (c) query formula, (d) objects, (e) suggested queries, and (f) visualisation, (**c**) InChorus [88], and (**d**) Orko [15] including (A) input box, (B) network visualisation, (C) access icons, (D) details container, (E) summary container, and (F) filter and visual encodings.

Moreover, all V-NLIs, including basic and advanced visualisations, used abstract graphical elements (lines, points, bars), though we did not encounter any use of symbolic graphical elements such as icons or glyphs.

We examined the **Visual Mapping Identification** in the previous works and found that 50% (10/20) of them only have *fixed* visual mapping [15,63,69,70,80,84,85,87,90,92]. Meanwhile, only 10% (2/20) of them support *user-defined* mapping [81,89], while 20% (4/20) use only the *rule-based* visual mappings [40,64,82,86] method. Finally, there are V-NLIs that support a combination of two visual mappings [39,83,88,91].

For example, Data@Hand [84] is a mobile application on which users can track their daily steps and sleep time, among others. It has basic fixed visualisations that are displayed when users open the application. On V-NLIs such as Orko [15], ConVisQA [69], Miva [90], and DataBreeze [85], when the dataset is uploaded, data are directly displayed with one predefined visualisation method. FlowNL [70] displays flow visualisations on a 3D world map with NL commands. GeCoAgent [80,87] both have fixed visualisations that are updated with NL queries. TransVis [63] and Valetto [92] automatically generate visualisations from natural language, though both of these systems have only one fixed visualisation. TransVis uses transient data (quality of service vs. time) that is visualised with a line area graph and Valetto uses a scatter plot to visualise tabular data (e.g., cars).

There are V-NLIs that allow user-defined visual mapping. For instance, Ava [81] and Iris [89] use NL to perform complex data science tasks such as statistical analysis and both support visualising data with one available visualisation when asked.

Snowy [40] is one of the V-NLIs that support rule-based visual mapping to select layouts and graphical elements. It has three visualisation methods—bar chart, scatter plot and line chart—and the system automatically selects and updates the visualisation method

depending on the user's queries and pre-defined visualisation mapping rules. They use an adaptation of the "Show Me" [37] algorithm to decide the visualisation method according to data attributes. They follow rules such as displaying a scatter plot if there are two quantitative attributes on the x and y axis or displaying a bar chart if there is one qualitative and one categorical attribute. Boomerang [82] and Talk2Data [64] use NL to provide users with multiple visualisations. In Boomerang, when the user asks a question [82], the system provides the user with various visualisation recommendations about the data, whereupon the user can further explore and ask more related questions. On the other hand, Boomerang uses the recommendation panel to display different visualisations with a combination of data attributes on scatter plots and bar charts. To show these recommendations, the system computes the degree of interest, relevance and timeliness of data attributes to show visualisations that are related to users' intents. To select which attributes to visualise, they compare their data and written text by transforming them into binary vectors.

Similarly, Talk2Data [64] generates multiple visualisations after NL queries and the system also annotates the visualisation and gives textual answers. The system follows a rule-based approach in which it associates different data facts with different visualisation methods. Data facts are extracted from the data. For example, a categorisation fact includes categorical data and is associated with a bar chart. Finally, Gamebot [86] helps users to analyse basketball and football games by giving textual information about the data and showing users related visualisations, assigning diverse statistical information about the game (e.g., individual player statistics) to different visualisation methods (e.g., game overview with a flow chart). Gamebot asks users if they are interested in visualisations and, to facilitate the analysis, it offers the users buttons to customise (when necessary) and then displays visualisations.

As we stated earlier, there are four V-NLIs that support a combination of several visual mapping strategies. For example, InChorus [88] uses rule-based mapping to automatically select a visualisation depending on the attribute type detected from users' queries and it also allows users to explicitly request a visualisation method. Onyx [91] uses fixed visualisation methods but users can change these methods using WIMP or NL. Moreover, Ref. [39] is the only V-NLI that has a combination of fixed and rule-based mapping. It has multiple fixed visualisations, although, if a user's query cannot be answered by existing visualisations, the system creates new appropriate visualisation using the aforementioned "Show Me" algorithm [37]. Finally, Chat2Vis is the only V-NLI that uses artificial intelligence (Large Language Models—LLM) for visual mapping. Moreover, users can specify in their query which type of chart they want to use to visualise the data.

Related to **View Transformation**, most of the explored works use a single view to visualise data (11/20, 55%) [15,21,69,70,81,85–89,91,92]. There are nine V-NLIs that have multiple views. Boomerang [82] displays multiple recommended visualisations simultaneously, while Talk2Data [64] generates multiple visualisations with annotations in a visualisation narrative style. In the case of MIVA [90], there are three fixed visualisations (bar, line, map), which are simultaneously updated to answer users' queries. Similarly, Evizeon [39] supports synchronised multiple views. Moreover, in Data@hand [84] and TransVis [63], multiple visualisations can be observed. Chat2Vis [83] demonstrates visualisation outputs using three views that use different LLM models to compare their performance. Finally, Orko [15] and FlowNL [70] have complementary visualisations in addition to primary ones.

Furthermore, there are V-NLIs that also support other view transformations. For example, Refs. [15,69,80,82,84] support Focus+Context. Data@hand's [84] users can analyse their sleep time across a month and they can ask the system to show the days the user woke up at 8 a.m. In this way, the system highlights the days the user woke up at 8 a.m. but also displays in the background in grey the data for the whole month. Similarly, Refs. [15,80] have network visualisations and users can highlight certain nodes to see in detail while viewing the whole visualisation in the background. ConVisQA [69] gives users the opportunity to see the whole hierarchy while highlighting certain parts in response to

users' NL queries. On the other hand, Boomerang [82] uses an approach similar to small multiples (i.e., grid-like layout) on the right-hand side of the screen as recommendations while letting users ask questions on the left-hand side, as well as displaying users charts in the chat window. Similarly, this can also be seen in Talk2Data [64], as users can observe multiple related visualisations at the same time. However, we did not encounter any V-NLIs with multi resolution among the selected articles.

4.3. Interaction Space

The **Interaction space** refers to all the interactions that users can make throughout the different stages of the visualisation pipeline (see Figure 9).
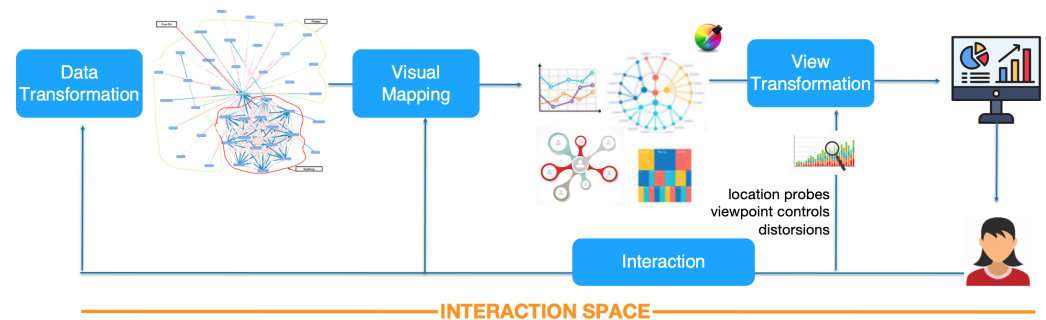

Figure 9. Interaction Space affects all the steps of the visualisation pipeline.

In the literature, the use of different interaction styles varies. Most V-NLIs (13/20, 65%) use both Basic (WIMP) and, naturally, Advanced (NL) interactions [15,39,40,63,69,70,84–86,88,90–92], while (7/20, 35%) of them use only Advanced (NL) interactions [64,80–83,87,89]. Additionally, in Table 3, we provide information on how each V-NLI used the seven interaction methods proposed by [10].

V-NLIs used the interaction techniques outlined in Table 3 at various stages of the pipeline illustrated in Figure 1. While some V-NLI interactions are designed for only one stage, others included interactions at multiple stages. The most used interaction techniques are select and filter. V-NLIs such as [63,80,82,84] use NL to interact with visualisations selecting (marking a data point) and filtering (showing something conditionally) data according to user queries. Boomerang [82] selects and filters data at the data transformation stage to create visualisations using NL. Data@hand [84] and TransVis [63] also use these techniques at the data transformation stage to update visualisations. Similarly, Refs. [87,92] use NL to update visualisations using filtering at the data transformation stage, while Chat2Vis [83] does this to generate visualisations. Others, such as [69,88,90,91], use both direct manipulation and NL to filter visualisations at the visual mapping stage. Refs. [39,40] use both basic and advanced interaction techniques at the visual mapping stage to filter visualisation and [39] uses advanced interaction while using the select method. Orko [15] and Databreeze [85] use both NL and direct manipulation to filter and select data on visualisations. Moreover, Gamebot [86] asks users if they want to see a visualisation related to their query, and before displaying the visualisation, the chatbot asks questions to users to filter the data to customise it before visualising and it gives users options with buttons. Similarly, Ava [81] uses NL to interact with data and not visualisations. It uses NL to perform complex data science tasks such as statistical analysis and generating visualisations from libraries. Finally, Ref. [64] uses advanced NLP-based interaction techniques when labelling selected data visualisation (maximum sale), and Ref. [70] uses both interaction styles.

The next most used method is Encode [15,40,63,64,83,85,88,89,91,92]. For example, Refs. [15,40,85,88,91] allow users to colour and size data points and add/remove attributes by using Basic (WIMP) and Advanced (NL) interactions at the visual mapping stage. On the other hand, Valetto [92] and TransVis [63] use NL commands to add or remove attributes at the data transformation stage. Similarly, Iris [89] uses NL to interact with data in Visual Mapping (i.e., users can select different attributes for axis), but not with View

Transformations. On the other hand, Talk2Data [64] and Chat2Vis [83] only interact at the Data Transformation stage, allowing NL queries to colour visualisations.

Table 3. Use of seven interaction methods in V-NLIs. N: Natural Language and W: WIMP—Windows Icons Menus Pointer.

V-NLI	Select	Explore	Reconfigure	Encode	Abstract/ Elaborate	Filter	Connect
ACUI [80]	N					N	N
Ava [81]	N						
Boomerang [82]	N					N	
Chat2Vis [83]				N		N	
ConVisQA [69]						W & N	
Data@Hand [84]	N				N	N	
DataBreeze [85]	W & N		W & N	W & N		W & N	
Evizeon [39]	N	W	N			W & N	
FlowNL [70]	W & N						
GameBot [86]						W	
GeCoAgent [87]						N	
InChorus [88]		W	W & N	W & N	W & N	W & N	
Iris [89]				N			
MIVA [90]						W & N	
ONYX [91]				W & N		W & N	
Orko [15]	W & N	W		W & N		W & N	N
Snowy [40]				W & N	N	W & N	
Talk2Data [64]	N			N			
TransVis [63]	N	W		N	W	N	
Valetto [92]			W	W & N		N	

The reconfigure method is supported by four V-NLIs [39,85,88,92], which are used to change the visual perspective of the data in the visual mapping. For instance, Valetto [92] uses gestures (a basic interaction) to flip the axis in the visualisation mapping stage. InChorus [88] uses both basic and advanced interaction methods, such as re-ordering data in the step of data transformation, to reconfigure the visualisation. Similarly, in the same step, Databreeze [85] uses a combination of basic and advanced interactions to rearrange data points and Evizeon [39] uses advanced interactions for this task.

Furthermore, the explore method, which is considered to be zooming and panning in the View Transformation stage, is used in four V-NLIs [15,39,63,88], all with basic interactions. It should be noted that Evizeon [39] and Orko [15] also automatically zoom in/out to the part of the visualisation that is related to users' query, though users cannot ask it to zoom in directly using NL. The abstract/elaborate method is used in four V-NLIs [40,63,84,88] to drill down to show more details. For example, Ref. [84] transforms data to show average hours of sleep over various months, and users can choose the visual mapping to see each month separately in more detail using NL. Similarly, Ref. [40] uses NL to do drill downs, while, on the other hand, TransVis [63] uses direct manipulation. InChorus [88] uses both modalities. Finally, the connect method is only used by two V-NLIs [15,80]. Both of these V-NLIs have network visualisation and use the connect method to highlight the relationships between links using Advanced interactions (i.e., using Focus+Context visualisations). While [66] performs this at the data transformation stage, Ref. [15] does this at the visual mapping stage.

4.4. Interactive Space of a V-NLI

Table 4 summarises the chatbot input categories in related work. Among the existing work, 50% (10/20) integrated **Chatbot-based V-NLIs** [63,80–83,86,87,89,91,92]. These V-NLIs have a chat window in which users can engage in conversations with a bot to analyse data visualisations. In some tools, the chat window is separated from the main visualisation dashboard [63,80,87,91,92], and in others, the visualisations are displayed in the chat windows [81–83,86,89]. For instance, both Iris [89] and Ava [81] were developed to help users perform complex data science tasks such as statistical analysis. While [89] displays visualisations in a single chat window, Ref. [81] has two windows, one containing the chatbot and the other showing the actions the chatbot performs, such as displaying visualisations. Moreover, we consider half of the approaches to be **Form-based V-NLIs** [15,39,40,64,69,70,84,85,88,90].

Table 4. Summary of input chatbot categories of V-NLIs. V-NLI interface (chatbot-based or form-based), Query Type (low or high), Follow-up query, Conversational Guidance: Help (data-based, user-based: based on what the user can ask), Auto-complete and Recommendation (recommend next action from D: Data, N: previous NL intent, W: previous WIMP interaction), and Input Modality.

V-NLI	V-NLI Interface		Input		Conversational Guidance			Multimodal.
	Chatbot	Form	Query T.	Follow-Up	Help	Autocom.	Recom.	
ACUI [80]	✓		low					
Ava [81]	✓		low		Hint/help		D, N	
Boomerang [82]	✓		low					
Chat2Vis [83]	✓		low & high					
ConVisQA [69]		✓	low			✓		WIMP
Data@Hand [84]		✓	low				D	Touch
DataBreeze [85]		✓	low	✓				Touch
Evizeon [39]		✓	low	✓		✓		WIMP
FlowNL [70]		✓	low			✓		WIMP
GameBot [86]	✓		low					WIMP
GeCoAgent [87]	✓		low			✓		
InChorus [88]		✓	low					Touch
Iris [89]	✓		low					
MIVA [90]		✓	low					WIMP
ONYX [91]	✓		low		data-based			WIMP
Orko [15]		✓	low	✓			N	Touch
Snowy [40]		✓	low	✓	data-based		D, N, W	WIMP
Talk2Data [64]		✓	low & high				D	
TransVis [63]	✓		low		user-based			WIMP
Valetto [92]	✓		low		user-based			Gestures

When we explored different **Query Types**, we found that most of the previous research presented V-NLIs that support only low-level queries (90% (18/20)) [15,39,40,63,69,70,80–82,84–92]. For instance, in Refs. [40,69,70,80,82,84,85,88,90–92], users can ask direct queries and receive answers such as filtered or highlighted data points on visualisations or new visualisations. Moreover, there are V-NLIs that have more specific datasets and the chatbot is designed to ask users questions or give prompts to perform the analysis [63,81,86,87,89]. For example, Ref. [86] asks users questions to show them visualisations about basketball or football games, and [87] does this to help users extract genomics data into tables. Moreover, Refs. [81,89] both ask users questions to complete data science tasks. Finally, two V-NLIs support both low and high level queries, Talk2Data, which is form-based [64] and Chat2Vis, which is chatbot-based [83]. Specifically, Talk2Data [64] uses high-level questions to interact with data using basic interaction techniques such as filtering, and they split high-level queries into smaller sub queries to find answers. An example from Talk2Data is, *"Which*

genre has more user reviews, fiction or non-fiction books?" They break down this question into two: *"How many reviews does the fiction book category have?"* and *"How many reviews does the non-fiction book category have?"*. On the other hand, Chat2Vis [83] is able to understand more complex queries such as *"Show the number of products with a price higher than 1000 or lower than 500 for each product name in a bar chart, and rank the y-axis in descending order?"* using several LLMs, which generate correct visualisations. Nevertheless, these models require some refinements because they may generate unnecessary extra information.

Additionally, these queries can be only One-turn or **Follow-up**. There are only four V-NLIs that support follow-up queries [15,39,40,85] and all of them support only low-level queries. After each query, Ref. [40] recommends follow-up queries on a list. On [15,39,85], users can refer to entities using determiners and pronouns.

One of the important characteristics of chatbots is having **Conversational Guidance**. In the visualisation context, chatbots can help users to ask the right questions, suggest possible queries, navigate them through visualisations, and explain the tool operations that chatbots can perform. According to the results, 40% (8/20) of the existing tools do not provide [80,82,83,85,86,88–90] the user with any conversational guidance, while the rest of the tools (12/20, five of them chatbot-based) recommend tasks or queries [15,40,64,81,84], help users [40,63,81,91,92], or auto-complete queries [39,69,70,87] designed to increase the discoverability of the NLI, helping users to understand what the NLI is capable of doing.

For example, users can ask for help from the chatbot in Valletto [92] and TransVis [63] regarding what users can ask the chatbot. Ava [81] gives hints on how to execute actions based on previous interactions. Onyx [91] helps with what it is able to do, and when something is not clear, it gives users instructions to go into the training interface and teach the system. Snowy [40] supports users by providing possible intents based on data before starting the analysis.

Moreover, Ava [81] gives users recommendations about how to continue the analysis, i.e., which actions it can do next. It also gives the users choices and asks them follow-up questions about whether they want to perform the action that the chatbot recommended. These recommendations are based on data and previous users' intents expressed in natural language. Data@Hand [84] and Talk2Data [64] recommend intents to users according to the data, and Orko [15] suggests to users possible operations on tool-tip when the system is not sure about a user's query. Snowy [40] offers three different kinds of recommendations. The first one are recommendations depending on the data, which are displayed at the beginning to start the analysis, since users may sometimes be new to the dataset and do not know what to ask. Moreover, it offers users recommendations as a follow-up intent depending on previous NL intents and WIMP interactions. Furthermore, some V-NLIs are designed to collect specific information from users in a structured format in which chatbots ask questions or give the users prompts to complete the analysis [81,86,87,89].

Finally, 13 of the reviewed V-NLIs have additional **Multimodality** to Natural Language (NL). For example, Refs. [39,80] have ambiguity widgets with which users can interact. Moreover, with V-NLIs [15,84,85,88], users can interact with the user interface using touch. Users can also select filters and interact with data without using NL. It should be noted that these systems have synchronised input modalities. For example, in [15], users can select a node with touch and ask a query about that node. Moreover, in [85], users can select data points and ask the system to move them to the left-hand corner.

Similarly, Refs. [40,90] have synchronised input modalities such as, when a user selects a part of the visualisation using the mouse while answering the query, the system remembers this selection. Refs. [70,91] have filters through which users can interact with them using WIMP. In TransVis [63], users can employ the WIMP to select a part of visualisation to explore in depth, while Gamebot [86] offers the users buttons during the conversation and Valetto [92] uses gestures to change visual encoding such as flipping the axis.

4.5. Chatbot Output

Finally, we explored the **Output** categories of the chatbot (see Table 5). Giving **Feedback** is one of the most important qualities of chatbots. All of the works in this review give the users textual feedback and some of them give visual feedback as well. The only exception is Chat2Vis [83], which, probably due to its recentness, is not yet integrated into a visualisation platform. Basically, textual feedback is used to inform or justify chatbot decisions to the users. Works such as [15,40,80,90] inform users about the success or failure of their queries. Moreover, Refs. [63,81,86,87,89] provide the users with informative feedback, additional explanations and follow-up questions to users to carry on the analysis. For example, after creating a decision tree, Ref. [81] can ask users if they want to see another plot. Refs. [63,89] ask users questions to continue the analysis, such as *"Which column should I use on the x-axis"* and *"What is the recovery time you want to use?"*

Table 5. Summary of output chatbot categories of V-NLIs.

V-NLI	Output	
	Feedback (Textual or Visual)	Int. Style (WIMP, VR, AR)
ACUI [80]	Textual (inform)	
Ava [81]	Textual (inform, additional explanation)	
Boomerang [82]	Textual (inform, additional explanation), Visual (Graph)	WIMP
Chat2Vis [83]	Visual (generating titles)	
ConVisQA [69]	Textual (inform, additional explanation, Visual (Changes in UI)	WIMP
Data@Hand [84]	Textual (inform), Visual (Changes in UI)	WIMP
DataBreeze [85]	Textual (inform), Visual (Changes in UI)	WIMP
Evizeon [39]	Textual (inform), Visual (Changes in UI)	WIMP
FlowNL [70]	Textual (to understand), Visual (Graph)	WIMP
GameBot [86]	Textual (inform, additional explanation) , Visual (Buttons)	WIMP
GeCoAgent [87]	Textual (inform, additional explanation)	
InChorus [88]	Textual (inform), Visual (Changes in UI)	WIMP
Iris [89]	Textual (inform, additional explanation)	
MIVA [90]	Textual (inform), Visual (Changes in UI)	WIMP
ONYX [91]	Textual (inform, additional explanation) , Visual (Changes in UI)	WIMP
Orko [15]	Speech (inform), Visual (Graph, Changes in UI)	WIMP
Snowy [40]	Textual (inform), Visual (Changes in UI)	WIMP
Talk2Data [64]	Textual (narrative, Visual (annotation)	WIMP
TransVis [63]	Textual (inform, additional explanation), Visual (Changes in UI)	WIMP
Valetto [92]	Textual (inform, additional explanation), Visual (Changes in UI)	WIMP

Works such as [15,39,84,85,88] proposed different informative feedback types. For example, Ref. [84] gives users three types of textual feedback: to confirm that it had applied the command to visualisation, to inform users that the command is not valid, and when it fails to understand. Databreeze [85] also has three available textual feedback types: to confirm successful action, after a follow-up command, and after partially understanding a command. Evizeon [39] has five types of textual feedback: (i) when the intent is understood and the result is shown, (ii) when it does not understand the request but the system guesses the nearest operable result, (iii) when the query is partially understood feedback appears with highlighting the unknown word, (iv) when it understands the query but cannot find any result, and (v) when it does not understand the intent. InChorus [88] has three different feedback styles, after a successful operation, after completing a successful operation but not having an effect on the visualisation (e.g., asking to sort by date but the data are already sorted by date), and after an invalid command. Orko [15] is the only one that gives informative feedback using speech and it supports giving feedback after successful and unsuccessful commands.

Moreover, Boomerang [82] informs users about the insights of the data and additionally gives answers to direct questions such as *"Is there a correlation between sales and profit?"*. Similarly, ConVisQA [69] gives answers to direct questions such as *"what is the most negative comment?"* while displaying the textual answer with updated visualisation. Although FlowNL [70] does not give users feedback to inform them, it asks users the meaning of words if it does not understand the given query. Moreover, ONYX [91] informs users about the action it has performed and gives instructions to users to teach the meaning of the unknown commands using WIMP. Valetto [92] provides feedback to inform users when there is a misunderstanding and provides additional information to users such as stating the correlation of two attributes. Finally, Ref. [64] provides explanations about visualisations for creating narrative storytelling.

Furthermore, we explored related work that provided users with additional visual feedback, such as supplementary graphs with main visualisation or changes on filters on the UI that have been applied by the chatbot. For example, Boomerang's [82] main goal is to show users multiple recommended visualisations related to users' queries on the right-hand side of the screen; however, relevant graphs are also displayed in the chat window when required. FlowNL [70] presents users with an ambiguity widget and has two auxiliary charts, one being a histogram displaying the velocity magnitude of hurricanes, while the other is a 2D map chart that is used to signal to specific regions. Additionally, visualisation is synchronised with a table. ConVisQA [69] visualises a hierarchical structure of comments on the main visualisation that is synchronised with actual comments displayed on the right side of the screen. Moreover, Orko [15] visualises additional charts (e.g., bar) and shows on the user interface whose filters are activated and display widgets in response to queries. Similarly, Ref. [39] presents related widgets after each query. Gamebot [86] displays buttons to assist the conversation.

V-NLIs such as [40,63,84,85,88,90–92] have visual feedback on the UI. For example, Ref. [84] displays an 'Undo' button after every query; further, the user interface changes according to queries such as displaying related filters. InChorus [88] and Snowy [40] show applied filters on the WIMP; additionally, in Snowy, selected attributes can be seen as well. Filters and attributes shown on the UI are updated after each query in [90,91]. Moreover, Valetto [92] highlights the recognised text in the chatbot's UI. For example, when a user asks to *"Add acceleration to the graph"*, it changes the colour of the 'Add' token in the user's sentence. Finally, Talk2Data [64] shows annotations with visualisations, and Chat2Vis [83] titles the visualisations from the users' query.

4.6. Technology behind V-NLIs

In this section, we briefly explore the software technologies used in the reviewed works. We can distinguish between those that directly use *NLP-toolkits* and those that use *chatbot frameworks*. For the former, we found multiple examples. The most used NLP-toolkit is open-source CoreNLP in Java [93]. For instance, Snowy [40], Miva [90] and Evizeon [39] all use it. Others use CoreNLP in combination with other toolkits, such as [15], which combines CoreNLP with NLTK [94] and AIML [95], and ConvisQA [69], which integrates CoreNLP with an ANTLR parser [96]. Some works use other NLP-toolkits; for example, Valetto [92] uses spaCy toolkit [97]. Finally, Data@Hand [84], which focuses on speech recognition, uses Apple speech framework [98] and Microsoft Cognitive Services [99] for IOS and Android devices, respectively, and Compromise NLP toolkit [100] to perform part-of-speech tagging. Among the V-NLIs that use chatbot frameworks, running independently from the visualisation module, we find: ACUI [80] using Rocket Chat open-source software [101]; Boomerang [82] based on IBM Watson Assistant [102]; GeCoAgent [87] based on Rasa [103]; and TransVis [63] employing Google Dialogflow4 [104].

Moreover, other works proposed customised solutions. Gamebot [86] uses rule-based word matching. Iris [89] uses domain-specific language that transforms Python functions into an automata (finite state machine). Ava [81] employs a state machine to control natural language conversations. FlowNL [70] uses a declarative language to filter and combine data

to derive structures and translates natural language queries into declarative specifications to render visualisations. Finally, the latest contributions to the field: Chat2Vis [83] uses LLMs, while Talk2Data [64] uses a novel decomposition model that is extended from sequence-to-sequence (deep neural networks) architectures.

5. Discussion

In the following, we review the research questions stated in Section 3.1 to explore how the use of chatbots may improve data visualisation and visual analysis, and also open up new research trends in this field.

5.1. RQ1: How Do Chatbot-Based V-NLIs Contribute to Interactions with the Data Space?

To answer this research question, we contrasted the results of the input characteristics of V-NLIs (Table 4) with how these systems deal with the data space stage in the visualisation pipeline (Table 1). We found that most of the works allow the users to express only low-level queries, and those that consider high-level queries do so with simple data types (see Figure 10, signal a and b) and attributes, i.e., tabular data with numerical and nominal attributes. Therefore, there is a gap in the use of natural language for the analysis of complex data (network and hierarchical) and also in the use of spatial and temporal attributes. This gap can be due to two reasons.
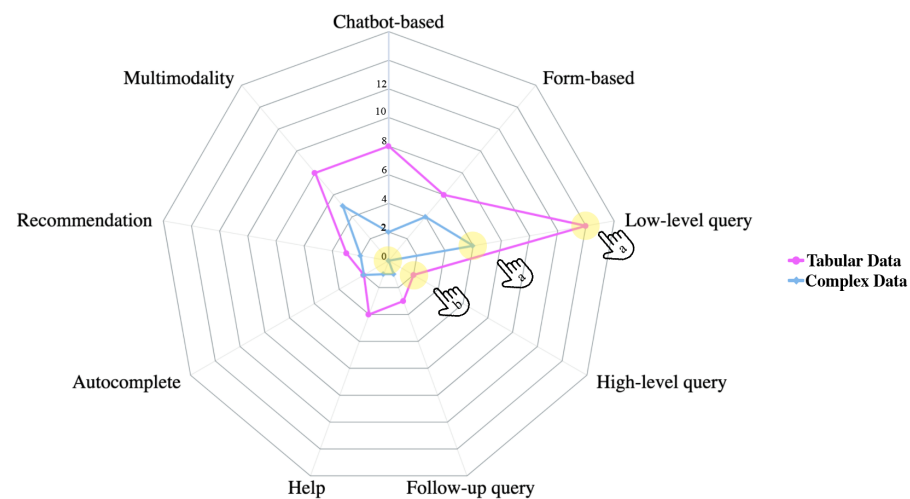

Figure 10. Spider chart displaying the relationship between data types and input V-NLI characteristics.

Remark 1. *We suggest designing V-NLIs considering complex data using high-level queries and extending their study to Post-WIMP interfaces, i.e., the so-called immersive analytics in VR and AR.*

First, low-level queries may make it difficult for users to perform visual analytic tasks with complex data (e.g., analysing subgraphs in network visualisations). Actually, the use of NLP to elaborate high-level queries on this type of data has limitations on both sides. On one hand, users need to express their intents. On the other hand, the NLP understanding system has to deal with ambiguities. Indeed, Talk2Data [64] and Chat2Vis [83] are the only reviewed works that used high-level queries, both with tabular data. However, the former has a form-based interface, and although the latter is chatbot-based, it lacks chatbot qualities such as conversational feedback and viewing conversation history. In this context, some recent approaches attempted to split NLI high-level intents directly into nested SQL-queries [105,106].

Second, complex data are usually projected into a two-dimensional space, hindering queries about complex structures, such as multivariate hierarchical and network data, which would be better queried in a three-dimensional space [107]. Therefore, we suggest designing V-NLIs considering high-level queries as well as extending their study beyond WIMP interfaces, i.e., the so-called immersive analytics in VR and AR [108].

Moreover, independently of the user's intents (low or high queries), all the examined V-NLIs contemplate simple data transformations (i.e., simple aggregations and statistical analysis such as correlations and logistic regressions). Note also that those simple data transformations have normally been incorporated into V-NLI systems that consider follow-up queries [39,40]. The reason can be found in analytical conversations, where this type of query makes it easier for the user to request successive data transformations beyond the initial or current visualisation. In this context, we suggest that V-NLI systems allow users to ask for more complex data transformations, such as visual binning or the extraction of subsets for the analysis of specific parts of the data [30,109,110] To do so, we think that combining Natural Language with other interaction styles may help the user to express the context of the visualisation, in line with the proposal of Beck et al. [63], where the user's NL-based queries refer to the part of the visualisation selected with the mouse. For example, suppose there is a 3D scene that shows a hierarchical graph. In this scenario, the user could utilise a VR hand controller to indicate the specific part of the visualisation to which the query refers. This idea would also be useful to indicate the target in focus+context and multi-view visualisations.

Remark 2. *Combining Natural Language with other interaction styles (VR, AR) may help the user to better express the data queries using the visual context during the conversation.*

Additionally, to let users better express their intents with less ambiguity, V-NLI systems use either guidance strategies or multimodality. Few works provide users with help or recommendations based on the data type, which is currently mainly tabular data [40,64,81,84]. We think that extending these guidance strategies to intricate data may improve human-chatbot interaction in terms of discoverability since the users can flow more directly through the visual analytics process based on those recommendations [111]. Regarding multimodality, most systems allow user–chatbot interaction combined with WIMP, but few of them allow touching [15,84,85,88] and only one work uses gestures [92]. Therefore, there are also opportunities for improvement particularly in relation to multimodality [112], which can also facilitate the transformation of the data since the users can communicate with the system in a more complete way (not only using text and voice but also gestures and gaze). Multimodality can also foster the development of a collaborative analysis of visualisations. Moreover, multimodality can be an additional input for the NLP system to enhance the context in analytical conversations.

Remark 3. *Multimodality can facilitate data transformations since the users can communicate with the system in a more complete way (not only using natural language but also gestures and gaze).*

5.2. RQ2: How Do Chatbot-Based V-NLIs Contribute to Interactions with the Visual Space?

By addressing this research question, we aim to shed light on how V-NLIs in the literature (Tables 4 and 5) can support users' tasks in the Visual Space (Table 2) of the visualisation pipeline. Figure 11 shows the scope of advanced and basic visualisations in both V-NLI and visualisation dimensions; see the borders in purple and green colour, respectively. As we can appreciate in the magenta- and blue-coloured polygons, V-NLIs that consider basic layouts embrace these dimensions in greater measure than those considering advanced layouts. Furthermore, the empty space of the spider reveals that there is a lot of room for research on different aspects of both basic and advanced visualisations in V-NLIs. This gap can be explained by the fact that the field is still in its early stages of development, and consequently, many researchers focused on exploring the foundational aspects of the technology. Moreover, the reviewed research works usually concentrated on one aspect of the V-NLI at a time. For example, some works investigated query recommendation [40,113,114], others explored multimodality [15,88], whereas others focused on designing personalised V-NLIs for specific data and user profiles such as data scientists [86,87].

It is especially interesting to focus the analysis on conversational guidance strategies (Auto-complete, Help, Recommendation and Follow-up; see Figure 11, purple arc) since they improve the interpretability/understanding of the NLI query, guide the user along the process of the analysis, and so have a positive impact on the whole user experience. During this review, we came across works that include different kinds of conversational guidance [15,39,40,63,64,69,81,84,85,87,91,92], few of them supporting multiple types [40,81]. Nevertheless, there is an unexplored aspect in these works, which is that guidance strategies can be designed by focusing on the visualisation pipeline. Indeed, this aspect allows us to analyse this RQ but in reverse: *"How the can visualisation process contribute to improve V-NLIs?"*. In this context, a recent research project proposes the so-called eXplainable NLI (XNLI) [115], which is based on a high-level grammar for statistical graphics (Vega-lite specification [116]). Thanks to this grammar, the system is able to provide the users with explanations of the following visualisation process as well as tips for interactively reviewing the natural language-based query. We firmly believe that this idea can be extended to more advanced visualisations thanks to recent proposals such as GoTree [117], a grammar that allows tree visualisations to be instantiated by specifying different aspects such as visual elements, layouts and coordinate systems.

Remark 4. *Chatbots' guidance strategies can be designed leaning on the visualisation pipeline. That is, "How can the knowledge about the visualisation process improve V-NLIs?"*
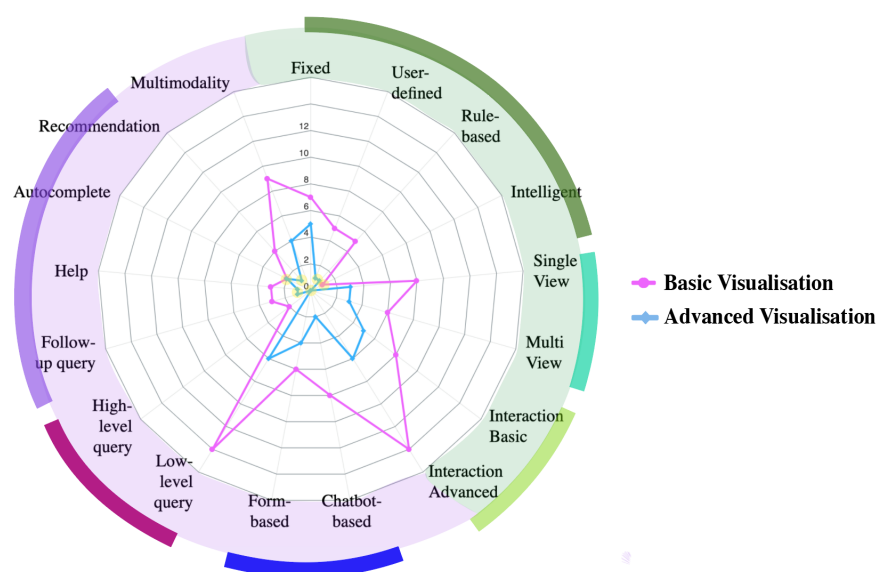

Figure 11. Spider chart displaying the relationship between Visual Space and V-NLI characteristics of analysed works.

Another way to facilitate a visual analysis, especially for inexperienced users, is to perform an automatic Visual Mapping, i.e., selecting the visualisation layouts and graphical elements automatically. When we explored related works, we found that most of the V-NLIs that support advanced visualisations do so with fixed layouts (see Figure 11, dark green arc). There is only one V-NLI that visualises an advanced visualisation (parallel plots) according to a rule-based visual mapping [88]. One possible reason for this lack of works may be that selecting visualisation layouts and graphical elements automatically is a complex task since it requires the V-NLI, first, to interpret user input accurately and, second, select the appropriate visualisation method based both on the data and on the context of the conversation. Moreover, most of the works that use rule-based visual mapping identification are form-based V-NLIs.

Note that only one study included in this scoping review explored intelligent Visual Mapping [83]. As a first step in this direction, DashBot [118] presents a new method

for training agents to imitate human exploration behaviour in visualisations using deep reinforcement learning. It has the potential to develop visualisation recommenders without requiring pre-existing training datasets. However, it uses simple data types (tabular) with basic visualisations and does not use NLP. Sevi [119] is another ML-based data visualisation system that creates visualisations using text or speech. Sevi's key component includes an end-to-end neural machine translation model called ncNet [28], which was evaluated using a cross-domain benchmark called nvBench [120]. The inputs of the model are an optional chart template and the NL query, outputting a chart styling of the rendered visualisation. Another approach is to combine user-defined Visual Mapping identification and automatic identification, which can give experienced users more freedom in their analysis, as demonstrated by Srinivasan et al. in their work with InChorus [88]. We think that the latter work paves the way to V-NLIs similar to those found in the field of mixed-initiative (human–machine collaboration) Procedural Content Generation [121].

Furthermore, recent advances in chatbot technology, such as ChatGPT-4 [122], demonstrate its ability to respond to visual queries. We firmly believe that these advances can also be applied to the field of data visualisation. For example, users can ask the chatbot to show them a visualisation of a particular layout by sending an image showing the desired layout. In fact, a recent study has focused on creating data visualisations using Natural Language with ChatGPT-3 and GPT-3.5 [83]. The study proposed using Large Language Models (LLMs) to create data visualisations from tabular data with basic visualisation methods. The system is able to select the appropriate visualisation type based on user queries. However, these advances come with several challenges, such as difficulties in specifying refinements to plotting elements, variability in the type of plot generated, and their non-deterministic nature. Given the fact that none of the works reviewed in this scoping review use NL interactions to change symbolic Graphical Elements, such as glyphs and colour palettes, these generative approaches can potentially be used to generate them during analytical conversations.

Remark 5. *Recent Generative AI models can potentially be used to generate visual layouts and graphical elements.*

Finally, regarding the spider graph in Figure 11 (see turquoise arc), we found that most of the V-NLIs that offer multiple views are form-based and have basic visualisations. Two of them have only an NL input modality [64,82]. However, others are multimodal, offering input modalities such as WIMP [39,63,70,90] and touch [15,84]. Additionally, some platforms allow the users to utilise two modalities simultaneously [15,85]. Although multimodality can be beneficial for any kind of visualisation, whether basic or advanced, we think that exploring multimodality, especially with advanced visualisations, is a promising research topic since they represent complementary inputs and outputs to improve the expressiveness of users' intentions and, consequently, the user experience in V-NLIs. For instance, users can ask to zoom in on a region of nodes or select a cluster that is on the side by showing gestures or, in a VR environment, using a VR controller to point. Alternatively, users could request a zoom level where the data are most densely clustered, or ask the system to identify the places of data points on the visualisation (e.g., by asking *"What's above the largest node?"* and then requesting further details). Additionally, graphic animations could be incorporated into the explanations provided by the system, enhancing the user's understanding of the data.

5.3. RQ3: How Do Chatbot-Based V-NLIs Enhance the User's Interaction with the Visualisation?

For this research question, we analysed the input (Table 4) and output (Table 5) characteristics of V-NLIs against the Interaction Space (seven interaction methods shown in Table 3). As can be appreciated in Figure 12, both chatbot-based and form-based approaches cover a similar, short range of interactive methods—Filtering and Selecting being the most covered—including some values near zero, especially with chatbot-based approaches

(see the complex interactions Abstract/Elaborate [63], Connect [15,80], Reconfigure [92], Explore [63] in yellow dots). This may be due to the difficulty of understanding when the user's intentions imply these complex interactions. Indeed, a recent study along these lines explores the use of a deep learning-based NL interpreter to translate NL utterances into editing actions, such as data operations (e.g., Filter, Aggregate), Encoding (e.g., changing colour, shape), and Reconfigure (e.g., position) [123]. Moreover, the emerging Large Language Models (LLM), which have proven their performance in various natural language tasks, open up new possibilities in Abstract and Elaborate interactions through step-by-step reasoning, such as the LLM Minerva developed by Google [124], which currently solves mathematical and scientific questions.
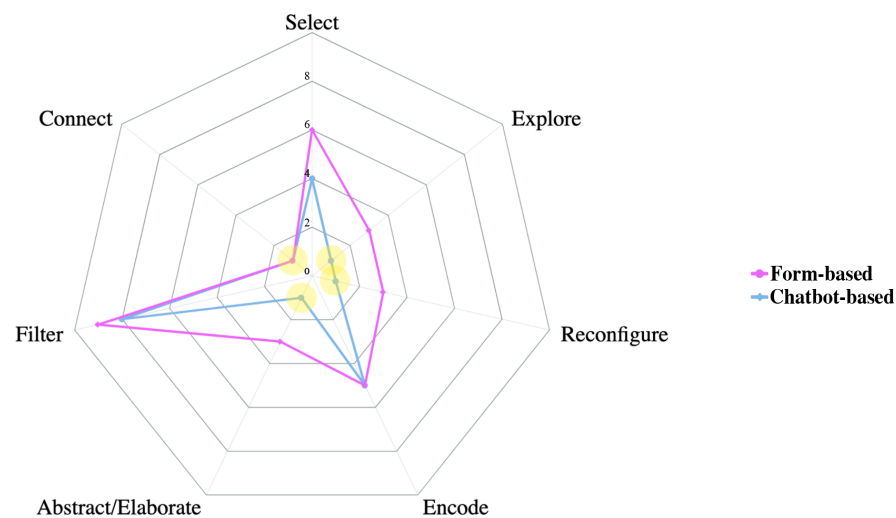

Figure 12. Spider chart displaying the relationship between the type of V-NLIs and interaction methods.

Another interesting finding is a passive listening mode that allows the chatbot to observe conversations happening between users and automatically proposes Select or Filter methods accordingly [80]. In line with this, a recent study explored an always-listening agent that acts as a third collaborator in a multi-person visual analysis. The agent generates visualisations based on observations it makes from users' conversations [125]. We think that this idea of passive listening can be extended with other input signals, such as eye tracking [126] and emotional measures such as the tone of voice [127].

Remark 6. *In collaborative scenarios, the chatbot may "observe" conversations happening between users and proactively propose the adequate interaction methods to perform users' tasks.*

Furthermore, although most V-NLIs support multiple input modalities (e.g., NL and WIMP or Touch), we did not encounter any V-NLI integrating VR and AR technologies. These technologies can easily provide additional inputs to the seven methods investigated, such as gaze, gestures and locations [128]. These technologies are not only important for input purposes, but also as additional means of enriching chatbot outputs, as in surround sounds, user's movements and haptic feedback using VR gloves or HMD (Head Mounted Displays). In fact, increasing the levels of immersion with multisensory stimulation has been demonstrated over the past decades to enhance data visual analysis tasks [108], although there is still room for improvement in terms of interactions with data visualisations. For instance, virtual teleportation is a common technique to guide users through data analysis in VREs. Teletransportation also could be used by the chatbot to situate the user near to the new generated visualisation that results from Select, Filter and Explore actions.

In fact, multisensory output systems can also be exploited by chatbots in non-immersive environments. In our scoping review, regarding the sound feedback, we found only one

incipient experiment with promising results that uses speech instead of textual output [15]. Indeed, there is a recent study [14] that compared voice vs. screen-based conversational agents created for purposes other than visualisation analysis. It observed that pairs of participants working together tended to take more conversational turns when speaking with a chatbot directly than when the same conversation is conducted in a chat window. However, in the specific context of visual analysis, both output systems (screen and sound) offer complementary advantages. That is, screen-based chatbots allow the users to track their conversation history, while sound-based chatbots allow them to seamlessly and quickly interact when working together. Thus, we suggest investigating how chatbots integrate both speech and textual conversations to support users' collaborations during the visual analysis.

In relation to complementary visual feedback, most of the reviewed V-NLIs provided the users with complementary visual feedback in the form of supplementary graphs and changes in the UI that provide information about chatbot responses. Nevertheless, there is still room for improvement. For instance, animated transitions [129] can help users to understand how changes in the visualisation settings are affecting the display and pop-up windows can show additional information or graphs. Additionally, the idea of visual narrative storytelling used in the reviewed form-based work [64] can be exploited in depth by chatbots helping users to summarise their data analysis findings. In line with this, it is important to take into account the lessons learned by the data analysis community during the last decade, such as the fact to avoid unbiased views of the explored data [130].

Remark 7. *Visual narrative storytelling can be exploited in depth by chatbots helping users to summarise their data analysis findings, guaranteeing unbiased views of the explored data.*

Additionally, in most of the reviewed chatbot works, textual feedback is used to inform users about the success or failure of their intents. Among them, there are works in which the V-NLIs also provide textual answers to direct questions [69,82], and short explanations about visualisation [64]. The current progress in generating LLMs definitely expands the scope of this kind of feedback, being able to provide more detailed explanations generated by LLMs with enriched information, such as including external links to detailed information of some topic. Moreover, text generation LLM from images offered by ChatGPT-4 [122] could be exploited by training it on the specific task of generating more information about the visualisations (i.e., transfer learning).

Remark 8. *The current progress in generating LLMs definitely expands the scope of chatbots' feedback, being able to provide more detailed explanations with enriched information.*

Last but not least, an important aspect in the development of any interactive system is the evaluation under the perspective of the Human–Computer Interaction (HCI) (i.e., ease of use, perceived usefulness, understanding and learnability, user satisfaction). Indeed, this aspect was not deeply covered in the V-NLI reviewed works, unlike that performed in other application domains (smartphone interfaces [131,132], web [133]).

6. Conclusions

This scoping review brings together the fields of data visualisation and chatbot-based interaction to study the body of literature on Visualisation-oriented Natural Language Interfaces (V-NLIs). Our aim is to provide an overall picture of the current state of V-NLIs and to identify and highlight future research directions. To do so, we first defined related categories and terminology for each space in the visualisation pipeline (Data Space, Visual Space, Interaction Space) and also outlined characteristics and key concepts of chatbots, following the proposed four dimensions (AINT—Anthropomorphic, Intelligent, Natural Language Processing, Interactive). Then, guided by three research questions that let us analyse prior V-NLIs with the lens of both fields, we provided a summary of the aspects that are currently focused on and supported by V-NLIs, as well as their limitations. Specifically,

the limitations found are related to the complexity of the analysed data, the type of queries supported by the chatbot, the lack of visual mapping automatisation, and the supported interaction styles. Finally, we highlighted and suggested potential promising research directions that may also help to overcome the aforementioned limitations. Specifically, exploring more advanced techniques in each dimension of the chatbot characterisation (AINT) will open up new challenges for the V-NLI research community.

References

1. Shen, H.; Bednarz, T.; Nguyen, H.; Feng, F.; Wyeld, T.; Hoek, P.J.; Lo, E.H. Information visualisation methods and techniques: State-of-the-art and future directions. *J. Ind. Inf. Integr.* **2019**, *16*, 100102. [CrossRef]
2. Kerren, A.; Purchase, H.C.; Ward, M.O. Introduction to multivariate network visualization. In *Multivariate Network Visualization*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–9.
3. Fu, S.; Wang, Y.; Yang, Y.; Bi, Q.; Guo, F.; Qu, H. Visforum: A visual analysis system for exploring user groups in online forums. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2018**, *8*, 1–21. [CrossRef]
4. Freeman, T.C.; Horsewell, S.; Patir, A.; Harling-Lee, J.; Regan, T.; Shih, B.B.; Prendergast, J.; Hume, D.A.; Angus, T. Graphia: A platform for the graph-based visualisation and analysis of high dimensional data. *PLoS Comput. Biol.* **2022**, *18*, e1010310. [CrossRef]
5. Zheng, B.; Sadlo, F. On the visualization of hierarchical multivariate data. In Proceedings of the 2021 IEEE 14th Pacific Visualization Symposium, Tianjin, China, 19–21 April 2021; pp. 136–145.
6. Stoiber, C.; Rind, A.; Grassinger, F.; Gutounig, R.; Goldgruber, E.; Sedlmair, M.; Emrich, Š.; Aigner, W. Netflower: Dynamic network visualization for data journalists. In Proceedings of the Computer Graphics Forum, Porto, Portugal, 3–7 June 2019; Volume 38, pp. 699–711.
7. Sondag, M.; Meulemans, W.; Schulz, C.; Verbeek, K.; Weiskopf, D.; Speckmann, B. Uncertainty treemaps. In Proceedings of the 2020 IEEE Pacific Visualization Symposium (PacificVis), Tianjin, China, 3–5 June 2020; pp. 111–120.
8. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009; Volume 3, pp. 361–362.
9. Shao, C.; Yang, Y.; Juneja, S.; GSeetharam, T. IoT data visualization for business intelligence in corporate finance. *Inf. Process. Manag.* **2022**, *59*, 102736. [CrossRef]
10. Yi, J.S.; ah Kang, Y.; Stasko, J.; Jacko, J.A. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1224–1231. [CrossRef]
11. Srinivasan, A.; Stasko, J. Natural language interfaces for data analysis with visualization: Considering what has and could be asked. In Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers, Barcelona, Spain, 12–16 June 2017; pp. 55–59.
12. Shen, L.; Shen, E.; Luo, Y.; Yang, X.; Hu, X.; Zhang, X.; Tai, Z.; Wang, J. Towards natural language interfaces for data visualization: A survey. *arXiv* **2021**, arXiv:2109.03506.
13. Belinkov, Y.; Glass, J. Analysis methods in neural language processing: A survey. *Trans. Assoc. Comput. Linguist.* **2019**, *7*, 49–72. [CrossRef]
14. Reicherts, L.; Rogers, Y.; Capra, L.; Wood, E.; Duong, T.D.; Sebire, N. It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. *ACM Trans. Comput.-Hum. Interact.* **2022**, *29*, 1–41. [CrossRef]
15. Srinivasan, A.; Stasko, J. Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 511–521. [CrossRef]
16. Murillo-Morales, T.; Miesenberger, K. Audial: A natural language interface to make statistical charts accessible to blind persons. In Proceedings of the Computers Helping People with Special Needs: 17th International Conference, ICCHP 2020, Lecco, Italy, 9–11 September 2020; Proceedings, Part I 17; Springer: Berlin/Heidelberg, Germany, 2020; pp. 373–384.

17. OpenAI. ChatGPT. 2021. Available online: https://openai.com/models/ (accessed on 8 May 2023).

18. OpenAI. DALL-E: A Generative Model for Diverse and Creative Images. 2021. Available online: https://openai.com/dall-e-2/ (accessed on 8 May 2023).

19. Srinivasan, A.; Stasko, J. How to ask what to say?: Strategies for evaluating natural language interfaces for data visualization. *IEEE Comput. Graph. Appl.* **2020**, *40*, 96–103. [CrossRef] [PubMed]

20. Cox, K.; Grinter, R.E.; Hibino, S.L.; Jagadeesan, L.J.; Mantilla, D. A multi-modal natural language interface to an information visualization environment. *Int. J. Speech Technol.* **2001**, *4*, 297–314. [CrossRef]

21. McTear, M. *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2020.

22. Hoque, E.; Kavehzadeh, P.; Masry, A. Chart Question Answering: State of the Art and Future Directions. *arXiv* **2022**, arXiv:2205.03966.

23. Card, M. *Readings in Information Visualization: Using Vision to Think*; Morgan Kaufmann: San Francisco, CA, USA, 1999.

24. Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*; Elsevier: San Francisco, CA, USA 2003; pp. 364–371.

25. Schulz, H.J.; Schumann, H. Visualizing graphs-a generalized view. In Proceedings of the Tenth International Conference on Information Visualisation (IV'06), London, UK, 5–7 July 2006; pp. 166–173.

26. Qin, X.; Luo, Y.; Tang, N.; Li, G. Making data visualization more efficient and effective: A survey. *VLDB J.* **2020**, *29*, 93–117. [CrossRef]

27. Hanrahan, P. Vizql: A language for query, analysis and visualization. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, Chicago, IL, USA, 27–29 June 2006; p. 721.

28. Luo, Y.; Tang, N.; Li, G.; Tang, J.; Chai, C.; Qin, X. Natural Language to visualization by neural machine translation. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 217–226. [CrossRef]

29. Satyanarayan, A.; Russell, R.; Hoffswell, J.; Heer, J. Reactive vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE Trans. Vis. Comput. Graph.* **2015**, *22*, 659–668. [CrossRef] [PubMed]

30. Kavaz, E.; Puig, A.; Rodríguez, I.; Chacón, R.; De-La-Paz, D.; Torralba, A.; Nofre, M.; Taule, M. Visualisation of hierarchical multivariate data: Categorisation and case study on hate speech. *Inf. Vis.* **2022**, *22*, 31–51. [CrossRef]

31. Börner, K.; Bueckle, A.; Ginda, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1857–1864. [CrossRef]

32. Khan, M.; Khan, S.S. Data and information visualization methods, and interactive mechanisms: A survey. *Int. J. Comput. Appl.* **2011**, *34*, 1–14.

33. Stolte, C.; Tang, D.; Hanrahan, P. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Trans. Vis. Comput. Graph.* **2002**, *8*, 52–65. [CrossRef]

34. Hoque, E.; Carenini, G. Convis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum*; John Wiley & Sons: Hoboken, NJ, USA, 2014; Volume 33, pp. 221–230.

35. Microsoft Power BI. 2023. Available online: https://powerbi.microsoft.com/es-es/ (accessed on 8 May 2023).

36. Chabot, C.; Stolte, C.; Hanrahan, P. Tableau software. *Tableau Softw.* **2003**, *6*. Available online: https://www.tableau.com/ (accessed on 8 May 2023).

37. Mackinlay, J.; Hanrahan, P.; Stolte, C. Show me: Automatic presentation for visual analysis. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1137–1144. [CrossRef]

38. Setlur, V.; Tory, M.; Djalali, A. Inferencing underspecified natural language utterances in visual analysis. In Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, CA, USA, 16–20 March 2019; pp. 40–51.

39. Hoque, E.; Setlur, V.; Tory, M.; Dykeman, I. Applying pragmatics principles for interaction with visual analytics. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 309–318. [CrossRef] [PubMed]

40. Srinivasan, A.; Setlur, V. Snowy: Recommending Utterances for Conversational Visual Analysis. In Proceedings of the The 34th Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, 10–14 October 2021; pp. 864–880.

41. Wang, C.; Feng, Y.; Bodik, R.; Dillig, I.; Cheung, A.; Ko, A.J. Falx: Synthesis-Powered Visualization Authoring. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021. [CrossRef]

42. Hu, K.; Bakker, M.A.; Li, S.; Kraska, T.; Hidalgo, C. VizML: A Machine Learning Approach to Visualization Recommendation. In Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI), Glasgow, Scotland, UK, 4–9 May 2019.

43. Borgo, R.; Kehrer, J.; Chung, D.H.; Maguire, E.; Laramee, R.S.; Hauser, H.; Ward, M.; Chen, M. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In Proceedings of the Eurographics (State of the Art Reports), Girona, Spain, 6–10 May 2013; pp. 39–63.

44. Zhu, S.; Sun, G.; Jiang, Q.; Zha, M.; Liang, R. A survey on automatic infographics and visualization recommendations. *Vis. Inform.* **2020**, *4*, 24–40. [CrossRef]

45. Yuan, L.P.; Zhou, Z.; Zhao, J.; Guo, Y.; Du, F.; Qu, H. Infocolorizer: Interactive recommendation of color palettes for infographics. *IEEE Trans. Vis. Comput. Graph.* **2021**, *28*, 4252–4266. [CrossRef]

46. Wang, Q.; Chen, Z.; Wang, Y.; Qu, H. Applying machine learning advances to data visualization: A survey on ml4vis. *arXiv* **2020**, arXiv:2012.00467.

47. Stasko, J.; Zhang, E. Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In Proceedings of the IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings, Salt Lake City, UT, USA, 9–10 October 2000; pp. 57–65.

48. Keim, D.A.; Schneidewind, J. Scalable visual data exploration of large data sets via multiresolution. *J. Univers. Comput. Sci.* **2005**, *11*, 1766–1779.

49. Dimara, E.; Perin, C. What is interaction for data visualization? *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 119–129. [CrossRef]

50. Amar, R.; Eagan, J.; Stasko, J. Low-level components of analytic activity in information visualization. In Proceedings of the IEEE Symposium on Information Visualization, 2005—INFOVIS 2005, Minneapolis, MN, USA, 23–25 October 2005; pp. 111–117.

51. Adamopoulou, E.; Moussiades, L. Chatbots: History, technology, and applications. *Mach. Learn. Appl.* **2020**, *2*, 100006. [CrossRef]

52. Kuhail, M.A.; Alturki, N.; Alramlawi, S.; Alhejori, K. Interacting with educational chatbots: A systematic review. *Educ. Inf. Technol.* **2022**, *28*, 973–1018. [CrossRef]

53. Bates, M. Health care chatbots are here to help. *IEEE Pulse* **2019**, *10*, 12–14. [CrossRef] [PubMed]

54. Thomas, N. An e-business chatbot using AIML and LSA. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 2740–2742.

55. Narechania, A.; Srinivasan, A.; Stasko, J. NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 369–379. [CrossRef] [PubMed]

56. Liu, C.; Han, Y.; Jiang, R.; Yuan, X. Advisor: Automatic visualization answer for natural-language question on tabular data. In Proceedings of the 2021 IEEE 14th Pacific Visualization Symposium (PacificVis), Tianjin, China, 19–21 April 2021; pp. 11–20.

57. Cassell, J. Embodied conversational interface agents. *Commun. ACM* **2000**, *43*, 70–78. [CrossRef]

58. Tellols, D.; Lopez-Sanchez, M.; Rodríguez, I.; Almajano, P.; Puig, A. Enhancing sentient embodied conversational agents with machine learning. *Pattern Recognit. Lett.* **2020**, *129*, 317–323. [CrossRef]

59. Ma, Z.; Dou, Z.; Zhu, Y.; Zhong, H.; Wen, J.R. One chatbot per person: Creating personalized chatbots based on implicit user profiles. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, 11–15 July 2021; pp. 555–564.

60. Khurana, D.; Koli, A.; Khatter, K.; Singh, S. Natural language processing: State of the art, current trends and challenges. *Multimed. Tools Appl.* **2022**, *82*, 3713–3744. [CrossRef] [PubMed]

61. Gatt, A.; Krahmer, E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *J. Artif. Intell. Res.* **2018**, *61*, 65–170. [CrossRef]

62. Gao, T.; Dontcheva, M.; Adar, E.; Liu, Z.; Karahalios, K.G. Datatone: Managing ambiguity in natural language interfaces for data visualization. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, NC, USA, 11–15 November 2015; pp. 489–500.

63. Beck, S.; Frank, S.; Hakamian, A.; Merino, L.; van Hoorn, A. TransVis: Using Visualizations and Chatbots for Supporting Transient Behavior in Microservice Systems. In Proceedings of the 2021 Working Conference on Software Visualization (VISSOFT), Luxembourg City, Luxembourg, 27–28 September 2021; pp. 65–75.

64. Shi, D.; Guo, Y.; Guo, M.; Wu, Y.; Chen, Q.; Cao, N. Talk2Data: High-Level Question Decomposition for Data-Oriented Question and Answering. *arXiv* **2021**, arXiv:2107.14420.

65. Keyvan, K.; Huang, J.X. How to Approach Ambiguous Queries in Conversational Search: A Survey of Techniques, Approaches, Tools, and Challenges. *ACM Comput. Surv.* **2022**, *55*, 1–40. [CrossRef]

66. Steinmetz, N.; Senthil-Kumar, B.; Sattler, K.U. Conversational Question Answering Using a Shift of Context. In Proceedings of the EDBT/ICDT Workshops, Nicosia, Cyprus, 23–26 March 2021.

67. Tory, M.; Setlur, V. Do what i mean, not what i say! design considerations for supporting intent and context in analytical conversation. In Proceedings of the 2019 IEEE Conference on Visual Analytics Science and Technology (VAST), Vancouver, BC, Canada, 20–25 October 2019; pp. 93–103.

68. Setlur, V.; Battersby, S.E.; Tory, M.; Gossweiler, R.; Chang, A.X. Eviza: A natural language interface for visual analysis. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, 16–19 October 2016; pp. 365–377.

69. Siddiqui, N.; Hoque, E. ConVisQA: A Natural Language Interface for Visually Exploring Online Conversations. In Proceedings of the 2020 24th International Conference Information Visualisation (IV), Melbourne, Australia, 7–11 September 2020; pp. 440–447.

70. Huang, J.; Xi, Y.; Hu, J.; Tao, J. FlowNL: Asking the Flow Data in Natural Languages. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 1200–1210. [CrossRef]

71. Hearst, M.; Tory, M. Would you like a chart with that? Incorporating visualizations into conversational interfaces. In Proceedings of the 2019 IEEE Visualization Conference (VIS), Vancouver, BC, Canada, 20–25 October 2019; pp. 1–5.

72. Otmazgin, S.; Cattan, A.; Goldberg, Y. F-COREF: Fast, Accurate and Easy to Use Coreference Resolution. *arXiv* **2022**, arXiv:2209.04280.

73. Huang, Y.; Wang, Y.; Tam, Y.C. Uniter-based situated coreference resolution with rich multimodal input. *arXiv* **2021**, arXiv:2112.03521.

74. Lee, B.; Riche, N.H.; Isenberg, P.; Carpendale, S. More than telling a story: Transforming data into visually shared stories. *IEEE Comput. Graph. Appl.* **2015**, *35*, 84–90. [CrossRef] [PubMed]

75. Segel, E.; Heer, J. Narrative visualization: Telling stories with data. *IEEE Trans. Vis. Comput. Graph.* **2010**, *16*, 1139–1148. [CrossRef]

76. Trani, J.F.; Browne, J.; Kett, M.; Bah, O.; Morlai, T.; Bailey, N.; Groce, N. Access to health care, reproductive health and disability: A large scale survey in Sierra Leone. *Soc. Sci. Med.* **2011**, *73*, 1477–1489. [CrossRef] [PubMed]

77. Peters, M.D.; Godfrey, C.M.; Khalil, H.; McInerney, P.; Parker, D.; Soares, C.B. Guidance for conducting systematic scoping reviews. *JBI Evid. Implement.* **2015**, *13*, 141–146. [CrossRef] [PubMed]

78. Tricco, A.; Lillie, E.; Zarin, W.; O'Brien, K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.; Horsley, T.; Weeks, L.; et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef]

79. Aromataris, E.; Riitano, D. Constructing a search strategy and searching for evidence. *Am. J. Nurs.* **2014**, *114*, 49–56. [CrossRef]

80. Bieliauskas, S.; Schreiber, A. A conversational user interface for software visualization. In Proceedings of the 2017 IEEE Working Conference on Software Visualization (Vissoft), Shanghai, China, 18–19 September 2017; pp. 139–143.

81. John, R.J.L.; Potti, N.; Patel, J.M. Ava: From Data to Insights through Conversations. In Proceedings of the CIDR, Chaminade, CA, USA, 8–11 January 2017.

82. Lee, D.J.L.; Quamar, A.; Kandogan, E.; Özcan, F. Boomerang: Proactive insight-based recommendations for guiding conversational data analysis. In Proceedings of the 2021 International Conference on Management of Data, Virtual Event, China, 20–25 June 2021; pp. 2750–2754.

83. Maddigan, P.; Susnjak, T. Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models. *arXiv* **2023**, arXiv:2302.02094.

84. Kim, Y.H.; Lee, B.; Srinivasan, A.; Choe, E.K. Data@ hand: Fostering visual exploration of personal data on smartphones leveraging speech and touch interaction. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–17.

85. Srinivasan, A.; Lee, B.; Stasko, J. Interweaving multimodal interaction with flexible unit visualizations for data exploration. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 3519–3533. [CrossRef]

86. Zhi, Q.; Metoyer, R. Gamebot: A visualization-augmented chatbot for sports game. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–7.

87. Crovari, P.; Pidò, S.; Pinoli, P.; Bernasconi, A.; Canakoglu, A.; Garzotto, F.; Ceri, S. GeCoAgent: A conversational agent for empowering genomic data extraction and analysis. *ACM Trans. Comput. Healthc. (HEALTH)* **2021**, *3*, 1–29. [CrossRef]

88. Srinivasan, A.; Lee, B.; Henry Riche, N.; Drucker, S.M.; Hinckley, K. InChorus: Designing consistent multimodal interactions for data visualization on tablet devices. In Proceedings of the 2020 CHI conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.

89. Fast, E.; Chen, B.; Mendelsohn, J.; Bassen, J.; Bernstein, M.S. Iris: A conversational agent for complex tasks. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.

90. Chowdhury, I.; Moeid, A.; Hoque, E.; Kabir, M.A.; Hossain, M.S.; Islam, M.M. MIVA: Multimodal interactions for facilitating visual analysis with multiple coordinated views. In Proceedings of the 2020 24th International Conference Information Visualisation (IV), Melbourne, Australia, 7–11 September 2020; pp. 714–717.

91. Ruoff, M.; Myers, B.A.; Maedche, A. ONYX-User Interfaces for Assisting in Interactive Task Learning for Natural Language Interfaces of Data Visualization Tools. In Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–7.

92. Kassel, J.F.; Rohs, M. Valletto: A multimodal interface for ubiquitous visual analytics. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–6.

93. Manning, C.D.; Surdeanu, M.; Bauer, J.; Finkel, J.R.; Bethard, S.; McClosky, D. The Stanford CoreNLP natural language processing toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, 23–24 June 2014; pp. 55–60.

94. Bird, S. NLTK: The natural language toolkit. In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, Sydney, Australia, 17–18 July 2006; pp. 69–72.

95. Bush, N.; Wallace, R.; Ringate, T.; Taylor, A.; Baer, J. Artificial Intelligence Markup Language (AIML) Version 1.0.1. Alice Foundation Work Draft. 2001. Available online: http://www.aiml.foundation/ (accessed on 8 May 2023).

96. Parr, T. *The Definitive ANTLR 4 Reference*; Torrossa: New York, NY, USA, 2013; pp. 1–326.

97. Honnibal, M.; Montani, I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *Appear* **2017**, *7*, 411–420.

98. AppleInc. Speech | Apple Developer. 2023. Available online: https://developer.apple.com/documentation/speech (accessed on 8 May 2023).

99. Microsoft. Cognitive Speech Services | Microsoft Azure. 2023. Available online: https://azure.microsoft.com/en-us/products/cognitive-services/speech-services/ (accessed on 8 May 2023).

100. Kelly, S. Compromise. 2019. Available online: http://compromise.cool/ (accessed on 8 May 2023).

101. Rocket.chat. 2023. Available online: https://www.rocket.chat/ (accessed on 8 May 2023).

102. IBM Watson Analytics. Available online: https://www.ibm.com/analytics (accessed on 8 May 2023).

103. Rasa. Rasa Conversational Platform. 2023. Available online: https://rasa.com/ (accessed on 8 May 2023).

104. Google Cloud. Dialogflow. 2023. Available online: https://cloud.google.com/dialogflow/ (accessed on 8 May 2023).

105. Sen, J.; Lei, C.; Quamar, A.; Özcan, F.; Efthymiou, V.; Dalmia, A.; Stager, G.; Mittal, A.; Saha, D.; Sankaranarayanan, K. Athena++ natural language querying for complex nested sql queries. *Proc. VLDB Endow.* **2020**, *13*, 2747–2759. [CrossRef]

106. Katsogiannis-Meimarakis, G.; Koutrika, G. A survey on deep learning approaches for text-to-SQL. *VLDB J.* **2023**, *32*, 905–936. [CrossRef]

107. Burch, M.; Vramulet, A.; Thieme, A.; Vorobiova, A.; Shehu, D.; Miulescu, M.; Farsadyar, M.; van Krieken, T. Vizwick: A multiperspective view of hierarchical data. In Proceedings of the 13th International Symposium on Visual Information Communication and Interaction, Eindhoven, The Netherlands, 8–10 December 2020; pp. 1–5.

108. Kraus, M.; Fuchs, J.; Sommer, B.; Klein, K.; Engelke, U.; Keim, D.; Schreiber, F. Immersive analytics with abstract 3D visualizations: A survey. In Proceedings of the Computer Graphics Forum, Virtual Event, OR, USA, 30 August–1 September 2022; Volume 41, pp. 201–229.

109. Cui, W.; Strazdins, G.; Wang, H. Visual Analysis of Multidimensional Big Data: A Scalable Lightweight Bundling Method for Parallel Coordinates. *IEEE Trans. Big Data* **2021**, *9*, 106–117. [CrossRef]

110. Moritz, D.; Wang, C.; Nelson, G.L.; Lin, H.; Smith, A.M.; Howe, B.; Heer, J. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 438–448. [CrossRef] [PubMed]

111. Setlur, V.; Hoque, E.; Kim, D.H.; Chang, A.X. Sneak pique: Exploring autocompletion as a data discovery scaffold for supporting visual analysis. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, 20–23 October 2020; pp. 966–978.

112. Saktheeswaran, A.; Srinivasan, A.; Stasko, J. Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 2168–2179. [CrossRef] [PubMed]

113. Shen, L.; Shen, E.; Tai, Z.; Xu, Y.; Dong, J.; Wang, J. Visual Data Analysis with Task-Based Recommendations. *Data Sci. Eng.* **2022**, *7*, 354–369. [CrossRef]

114. Wang, X.; Cheng, F.; Wang, Y.; Xu, K.; Long, J.; Lu, H.; Qu, H. Interactive data analysis with next-step natural language query recommendation. *arXiv* **2022**, arXiv:2201.04868.

115. Feng, Y.; Wang, X.; Pan, B.; Wong, K.K.; Ren, Y.; Liu, S.; Yan, Z.; Ma, Y.; Qu, H.; Chen, W. XNLI: Explaining and Diagnosing NLI-based Visual Data Analysis. *IEEE Trans. Vis. Comput. Graph.* **2023**. [CrossRef] [PubMed]

116. Satyanarayan, A.; Moritz, D.; Wongsuphasawat, K.; Heer, J. Vega-lite: A grammar of interactive graphics. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 341–350. [CrossRef]

117. Li, G.; Tian, M.; Xu, Q.; McGuffin, M.J.; Yuan, X. Gotree: A grammar of tree visualizations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.

118. Deng, D.; Wu, A.; Qu, H.; Wu, Y. Dashbot: Insight-driven dashboard generation based on deep reinforcement learning. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 690–700. [CrossRef]

119. Tang, J.; Luo, Y.; Ouzzani, M.; Li, G.; Chen, H. Sevi: Speech-to-visualization through neural machine translation. In Proceedings of the 2022 International Conference on Management of Data, Philadelphia, PA, USA, 12–17 June 2022; pp. 2353–2356.

120. Luo, Y.; Tang, J.; Li, G. nvBench: A Large-Scale Synthesized Dataset for Cross-Domain Natural Language to Visualization Task. *arXiv* **2021**, arXiv:2112.12926.

121. Shaker, N.; Togelius, J.; Nelson, M.J.; Liapis, A.; Smith, G.; Shaker, N. Mixed-initiative content creation. In *Procedural Content Generation in Games*; Springer: Cham, Switzerland, 2016; pp. 195–214.

122. OpenAI. GPT-4 Technical Report. *arXiv* **2023**, arXiv:cs.CL/2303.08774.

123. Wang, Y.; Hou, Z.; Shen, L.; Wu, T.; Wang, J.; Huang, H.; Zhang, H.; Zhang, D. Towards Natural Language-Based Visualization Authoring. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 1222–1232. [CrossRef] [PubMed]

124. Lewkowycz, A.; Andreassen, A.; Dohan, D.; Dyer, E.; Michalewski, H.; Ramasesh, V.; Slone, A.; Anil, C.; Schlag, I.; Gutman-Solo, T.; et al. Solving Quantitative Reasoning Problems with Language Models. *arXiv* **2022**, arXiv:cs.CL/2206.14858.

125. Tabalba, R.; Kirshenbaum, N.; Leigh, J.; Bhatacharya, A.; Johnson, A.; Grosso, V.; Di Eugenio, B.; Zellner, M. Articulate+: An Always-Listening Natural Language Interface for Creating Data Visualizations. In Proceedings of the 4th Conference on Conversational User Interfaces, Glasgow, UK, 26–28 July 2022; pp. 1–6.

126. Li, F.; Lee, C.H.; Feng, S.; Trappey, A.; Gilani, F. Prospective on eye-tracking-based studies in immersive virtual reality. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; pp. 861–866.

127. Amazon Web Services Inc. Amazon Lex Documentation. 2023. Available online: https://docs.aws.amazon.com/lex/index.html (accessed on 8 May 2023).

128. Zhao, Y.; Jiang, J.; Chen, Y.; Liu, R.; Yang, Y.; Xue, X.; Chen, S. Metaverse: Perspectives from graphics, interactions and visualization. *Vis. Inform.* **2022**, *6*, 56–67. [CrossRef]

129. Pulo, K. Navani: Navigating large-scale visualisations with animated transitions. In Proceedings of the 2007 11th International Conference Information Visualization (IV'07), Zurich, Switzerland, 4–6 July 2007; pp. 271–276.

130. Vasconcelos Braga, J.; Silva, T.B.P.e. Storytelling in data visualization: Information bias. *InfoDesign-Rev. Bras. Des. Inf.* **2021**, *18*, 53–66. [CrossRef]

131. Khan, I.; Khusro, S. ConTEXT: Context-aware adaptive SMS client for drivers to reduce risky driving behaviors. *Soft Comput.* **2022**, *26*, 7623–7640. [CrossRef]

132. Khan, I.; Khusro, S. Towards the design of context-aware adaptive user interfaces to minimize drivers' distractions. *Mob. Inf. Syst.* **2020**, *2020*, 8858886. [CrossRef]
133. Insfran, E.; Fernandez, A. A systematic review of usability evaluation in web development. In Proceedings of the Web Information Systems Engineering—WISE 2008 Workshops: WISE 2008 International Workshops, Auckland, New Zealand, 1–4 September 2008; Proceedings 9; Springer: Berlin/Heidelberg, Germany, 2008; pp. 81–91.