A Pilot Study Comparing ChatGPT and Google Search in Supporting Visualization Insight Discovery

He, Chen; Welsch, Robin; Jacucci, Giulio

Soto, Axel; Zangerle, Eva

2024

http://hdl.handle.net/10138/575135

He, C, Welsch, R & Jacucci, G 2024, A Pilot Study Comparing ChatGPT and Google Search in Supporting Visualization Insight Discovery. in A Soto & E Zangerle (eds), Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA. CEUR Workshop Proceedings, vol. 3660, CEUR-WS.org, Aachen, Workshops at the International Conference on Intelligent User Interfaces, Greenville, South Carolina, United States, 18/03/2024.

A Pilot Study Comparing ChatGPT and Google Search in Supporting Visualization Insight Discovery

Chen He[1,*], Robin Welsch[2] and Giulio Jacucci[1]

[1]University of Helsinki, Finland
[2]Aalto University, Finland

Abstract

The popularity of large language models (LLMs) provides new possibilities for deriving visualization insights, integrating human and machine intelligence. However, we have yet to understand how a contextualized LLM compares with the traditional search in supporting visualization insight discovery. To this end, we conducted a between-subjects study with 25 participants to compare user insight generation with chat/search on a CO_2 Explorer. The Chat condition has ChatGPT contextualized with the data, user tasks, and interactions as programmed system prompts. Results show both systems have their merits and demerits: ChatGPT affords users to ask more diverse questions but can produce wrong answers; Search provides information sources, making the answer more reliable, but users can fail to find the answer. This study prompts us to synthesize them in a future study for reliable and efficient information retrieval.

Keywords

Information Visualization, Large Language Models, Google Search, Empirical Study

1. Introduction

Discovering insights is considered the main purpose of visual data exploration (VDE) [1]. Compared with data tables, visualization reveals data patterns and trends, facilitating insight discovery. But still, deriving insights needs visualization literacy and cognitive efforts [2]. Imagine that an AI system could provide insights into the data you are exploring right now instead of you meticulously looking for them. Prior work proposed techniques to (semi-)automate insights, such as data trends and clusters (e.g., [3, 4, 5]); however, researchers pointed out the superficiality of automated insights: 1) Automated insights are limited to the data while losing the context of the domain. For instance, automatically discovered data clusters and patterns might not be meaningful to the domain under exploration to improve the viewer's understanding of the domain [6]. 2) Deriving knowledge from collected insights could not be automated [7]. Analysts often need to gather evidence from multiple perspectives to build new knowledge. However, the advent of the Large Language Models (LLMs) might update our views on how visualization insight could be generated.

Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA
*Corresponding author.
chen.he@helsinki.fi (C. He); robin.welsch@aalto.fi (R. Welsch); giulio.jacucci@helsinki.fi (G. Jacucci)
0000-0003-2055-4468 (C. He); 0000-0002-7255-7890 (R. Welsch); 0000-0002-9185-7928 (G. Jacucci)

The introduction of ChatGPT popularized LLMs thanks to its advantages in a wide range of tasks and simple conversational interfaces, despite its limitations like hallucinations [8]. LLMs can benefit visualization in multiple ways, such as charts' title generation and pattern recognition [9]. LLMs' large reservoir of information and the modeling thereof provide the potential to level up visualization insight generation, including but not limited to linking data to external evidence to build a plausible insight and using its reasoning capability to derive hypotheses and generalizations.

This research investigates the use of contextualized LLMs (ChatGPT 4 with Vision [10]) to facilitate visualization insight generation, compared with using the traditional Google search interface to answer the research question (RQ): **What are the similarities and differences between contextualized ChatGPT and Google search in supporting visualization insight generation?**

We provide ChatGPT with contextual information by prompting the data under exploration, user tasks, and real-time user interactions and the resulting visualization states as system messages. To investigate the RQ, we conducted a between-subjects study with 25 participants, asking them to explore an existing CO_2 Explorer and discover data insights with external evidence focusing on either insight quantity or quality. The CO_2 Explorer has a chat/search interface next to the visualization as two conditions for comparison. Results show that both have their own strengths and weaknesses, which prompt us to integrate them for future studies.

2. Related Work

2.1. Natural Language Support for Data Insights

Manually generating data insights can be time-consuming and opportunistic; prior work developed techniques to discover data insights, like averages and extremes, systematically and generate texts and visualizations to communicate the insights (e.g., [11, 12]). The process can be conversational in question-and-answer mode: The user queries the data in natural language, and the system provides textual and/or visual answers (e.g., [13, 14, 15]). On the other hand, prior research explored computationally linking the visualization and its textual annotations for visual storytelling/presentation (e.g., [16, 17]). In contrast, our study explores how users use contextualized LLMs to generate visualization insights compared with using Google search. The work most close to ours, to the best of our knowledge, is DataTales [18], which asks users to author data stories using LLMs. The generated narrative is linked to the chart components. However, DataTales does not use the conversational feature of LLMs but generates a narrative with predetermined prompts and was not compared with a baseline system.

2.2. ChatGPT vs. Google Search

Researchers compared ChatGPT and Google search in supporting medical information retrieval [19, 20, 21] and learning [22]. Results show that readability is low for both platforms [20], indicating that the information provided is not easy to understand by the general audiences, but ChatGPT is more difficult to read and comprehend [20, 19]. ChatGPT provides more relevant responses without citing sources, while Google is more reliable as it often attaches the date and

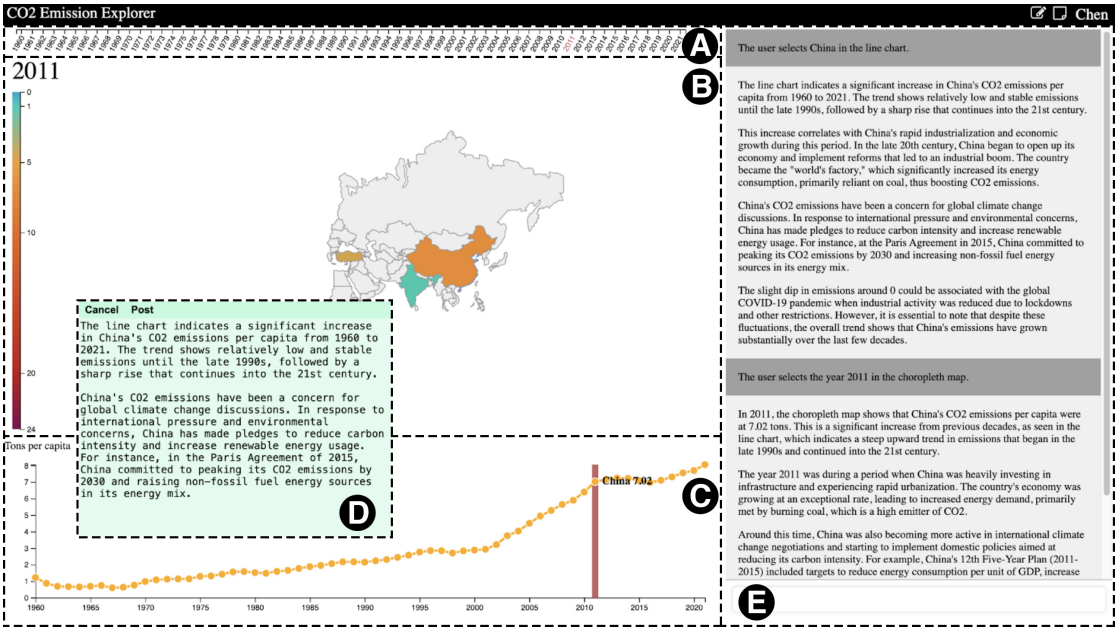

Figure 1: Screenshot of the interface of ChatGPT-empowered CO_2 Explorer for insight discovery. Users can select a year from the top list (A) to view that year's CO_2 emission of various countries on the map (B), select countries from the map to view their historic CO_2 emission in the line chart (C), chat with the chatbot (E) to gain more information about the data, such as news and events, and compose a note recording their discoveries (D).

source of retrieved information [20]. Ayoub et al. [21] found that GPT is better at providing general medical information but worse at medical recommendations compared with Google search. When solving programming exercises, Elissa and Marco [22] discovered that students using ChatGPT have a better success rate with less time spent but worse at their understanding of the topic when tested with questionnaires. We compare these two platforms in supporting visualization insight generation and draw conclusions to compare and contrast with prior results.

3. Study Design

To investigate the RQ, we conducted a between-subjects study comparing ChatGPT and Google search in assisting VDE. We developed a prototype integrating a CO2 Explorer and the chatbot/search. The CO2 Explorer, studied in prior work [23, 24], shows various countries' CO2 emission data in tons per capita from 1960 to 2021. It consists of a choropleth map and a line chart (Figure 1). Users can select a year from the top list to view that year's CO2 emission of various countries on the choropleth map (Figure 1B); mousing over a country on the map displays a tooltip with the country name and its emission value. Users can also select countries from the map to view these countries' histories of CO2 emission in the line chart (Figure 1C). Mousing over the line chart displays a black vertical reference line marking the year nearest to

the mouse pointer and that year's emission values of selected countries. The red vertical line in the line chart indicates the year chosen in the map view. To capture user insights during their VDE, users can input and post written texts as notes (Figure 1D).

3.1. ChatGPT to Assist Visualization Insight Discovery

We added the chat function to the visualization's right side to assist users in VDE. With ChatGPT 4 with Vision API (parameter settings in Appendix A), we feed in two types of contextual information as system prompts: 1) Description of the situation the user is in. The initial system prompt conveys the data in CSV format, describes the visualization and the user task, and instructs the chatbot to assist with the user task (Appendix B1). 2) To assist real-time insight discovery, the Explorer transforms user interactions as prompts to retrieve relevant information. The Explorer prompts three types of user interactions: user selection of a year and selection/de-selection of a country from the line chart, with the text describing the user interaction (Appendix B2) and the resulting visualization as an image prompt. In the study, users were unaware of the initial system message or the image prompt of their interaction, but they can see their interaction as a textual prompt and the response from the chatbot (Figure 1). So, every time they click a country or year, they need to wait for the answer to complete until they can click another one. Users can also prompt freely using the input box at the bottom right.

3.2. Baseline with Search Engine

As a baseline, we put a search component powered by Google Search API [25] next to the visualization instead of a chat interface. So users can use the search engine to assist with insight discovery. Unlike the chatbot, the search engine does not have information about the visualization or user interaction.

3.3. Participants

We recruit 25 international students from a large university to join the study through mailing lists. They conducted the study either on-site or remotely through Zoom. Upon completion, each received a 10-euro plus a bonus gift card from a local supermarket chain. They were randomly assigned to one of the two study conditions. Of the 12 in the Search condition (age range: 21-53, median: 25.5; female: 10; on-site: 4), five originally came from Asia, four were from Europe, and three were from North America. Of the 13 people in the Chat condition (age range: 21-45, median: 25; female: 6; on-site: 6), one was from North America, while the remaining six each were from Asia and Europe.

Except two from the Search condition and one from the Chat condition had an intermediate level of English proficiency, others self-indicated as having a native/advanced level of English. We examined their familiarity with the techniques in 5-point Likert scales. Both groups were familiar with heatmaps (Median search: 5, chat: 4) and line charts (Median search: 5, chat: 5). However, with the two test conditions, a Wilcoxon rank-sum test shows a statistically significant difference that the search group is familiar with Google search (median: 5) while the chat group (median: 3) is not so familiar with ChatGPT (Wilcoxon effect size: 0.76, $p < 0.001$).

3.4. Procedure and Tasks

The study consists of three stages: an interactive tutorial, two visual exploration tasks, and a questionnaire. The tutorial, built on top of the interface using the intro.js library [26], introduced the charts, note posting, and chat/search component in six steps. To capture user behavior in generating different types of insights for comparison, we created two tasks: one quantitative and one qualitative task. The task description of the **quantitative** task is:

> Freely explore the CO2 emission data of [a Country Group][a].
>
> *Post **as many notes as possible**, recording your data discoveries. Your data discoveries must be **linked to external evidence** as references, such as events, policies, and news.*
>
> Please use the Search function (or 'chat with the ChatBot' for the Chat condition) to assist with your task.
>
> *You will receive a maximum bonus of €5 for this task based on **the number of correct notes** you have posted.*
>
> [a]There are two country groups: 1. the USA, Italy, and Finland; 2. China, India, and Turkey

For the **qualitative** task, we replace the above italic part of the description with:

Post **one note** with the following requirements:

1. The note records **a hypothesis or generalization** you have made from your data analysis;
2. The note includes the rationale behind, that is, **how you have derived the hypothesis or generalization**;
3. The rationale **must** link your data analysis with **external evidence** as references, such as events, policies, and news;
4. The note **must be logical and correct**.

You will receive **a maximum bonus of €5** for this task if your note **satisfies the above requirements**. If you post multiple notes, only the last one will be evaluated.

The order of the two country groups and two tasks were randomly assigned to control the carryover effect. Finally, the questionnaire collected subjects' backgrounds, system usability scale (SUS [27]) answers, and free-form comments. Participants went through the whole study at their own pace. The experimenter was present if they had any questions. They were encouraged to think aloud during the tasks; we recorded the screen and voice for analysis. The whole study generally took less than an hour.

3.5. Data Collection and Analysis

We recorded the screen and voice during the tasks. Mouse interactions, user notes, query/prompt input, and search results/chat answers were logged with time stamps. Mouse hover actions were recorded if they lasted for over 3 seconds. We analyzed the time they spent on the tasks, the number of notes for the quantity task, the number of VDE actions, and their questionnaire answers. We assessed the overall gradings of notes on a 5-point Likert scale (grading criteria in Table 1). One author went through the notes, created the grading criteria, and graded the rest of

Table 1
Note grading criteria.

Grade	Interpretation
5	Better than Grade 4 with novel hypothesis/generalization.
4	Clear hypothesis/generalization; well-thought rationale; the logic makes the texts flow well.
3	Well-thought discovery of the data with multi-aspect evidence.
2	Simple discovery of the data with external evidence.
1	Unclear note, missing data references or external evidence.

the notes. For the quantity task, we averaged the note grades of each participant for statistical analysis, while for the quality task, each had one note, so there was no need for averaging.

Since two users in the Chat condition spontaneously used search engines, we removed them from the analysis except for the questionnaire part. For statistical analysis, we used the Wilcoxon rank sum tests (for unpaired samples) to compare performance between the two conditions; also, we used Wilcoxon signed-rank tests (for paired samples) to compare the two tasks within a condition. We report Wilcoxon effect sizes and p-values of the tests. Moreover, we examined participants' queries/prompts and went through video recordings to understand several action patterns when participants posted notes.

4. Results

Figure 2 shows that users spent more time on the quantity task than the quality task (Search effect size = 0.30, $p = 0.38$; Chat effect size = 0.69, $p = 0.03$), while the same task took a similar amount of time across the conditions. The number of notes recorded in the quantity task is also alike in both conditions (Median search: 4.5, chat: 3.5; Range search: [1, 11], chat: [1, 12]).

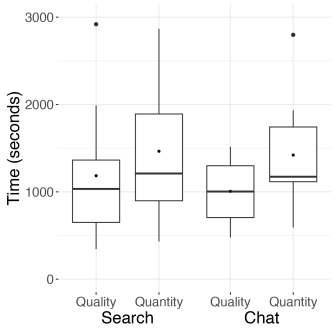

Figure 2: Time participants spent on the tasks.

In the questionnaire, we asked about their confidence in the notes they posted and if they learned something new during the tasks on 5-point Likert scales. While the two conditions showed no difference in user confidence in their notes (Median search: 4, chat: 4), the Chat condition demonstrated moderately more learning experiences for the users (Median search: 4.5, chat: 5; effect size = 0.34, $p = 0.11$).

The SUS scores reveal that the Search condition is considered slightly more user-friendly (Median search: 85, chat: 77.5; effect size = 0.27, $p = 0.18$), while both conditions are rated above the general average score of 68. User comments on the Chat condition show that five users complained about the VDE that they needed to wait for the chat answers after clicks; two users mentioned trust issues in the answers from ChatGPT; two had difficulties writing prompts to get the expected answers. Four pointed out that the chatbot gave pertinent answers (context-aware). With search, most comments are about the visualization instead, while one mentioned the search results seemed repetitive.

To understand more about VDE in the two conditions, we counted the number of effective clicks and mouseovers on the charts. Results show that the Search condition had more data exploration actions than the Chat condition (Figure 3), especially for the quality task (Quality effect size = 0.59, p = 0.01; Quantity effect size = 0.28, p = 0.22). We can presume that had the Chat condition not blocked the user interaction, participants would have interacted more with the charts. In both conditions, the quantity task had more data exploration actions than the quality task (Search effect size = 0.47, p = 0.16; Chat effect size = 0.73, p = 0.02).

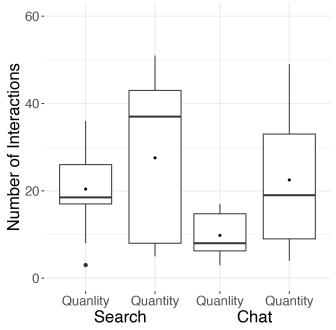

Figure 3: Number of VDE actions.

The quality task produced notes with significantly higher grades than the quantity task within the conditions with large effect sizes (Search effect size = 0.55, p = 0.08; Chat effect size = 0.81, p = 0.02). On the other hand, the two conditions did not show much difference in note grades with the two tasks (Figure 4).

Five among the 12 people (42%) in the Search condition put external links as evidence in notes in both tasks. Two users in the Chat condition who used search also included external links in notes. Some users made notes without mentioning year or country, such as using the phrase 'this is..', supposing the note is linked to the visualization state.

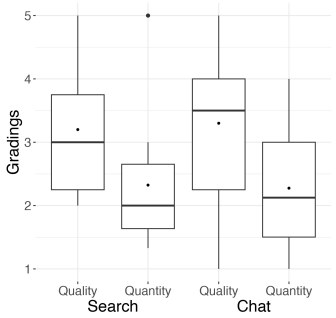

Figure 4: Note gradings.

Video analysis shows that in both conditions, almost all users pasted texts from websites/chats to notes. In the Chat condition, one user copied large amounts of texts as notes without reading the chats; another user put a wrong answer from chats directly as a note (an answer to which year has the biggest decrease in CO2 emission). In the Search condition, three out of 12 users failed to find the answer they were looking for. The Search condition allowed users to explore many external charts, infographics, and scientific articles.

Queries in search and prompts in chats showed similar qualities. Both are iterative; users often drill down to retrieve more concrete information. Both asked for facts, like events and policies, and causalities, such as the impact of renewable energy. However, with chats, questions are more diverse, including how much-, how-, and when-type of questions, such as "How much in absolute terms did the emissions of China go up from 2002 to 2011?" and "When can we say that the Kyoto protocol had a sure effect on the decrease on the emissions?". Moreover, queries are most often phrases, while prompts are complete questions.

5. Discussion

To summarize, results showed no significant differences between the two conditions in the time taken or the grades of notes for the tasks; neither did the number of notes generated for the quantity task. The result could be partially caused by participants' unfamiliarity with

the new ChatGPT technology. In both conditions, the quantity task took more time than the quality task, while the quality task produced notes with higher grades. In both conditions, users were confident in the notes they posted, while the Chat condition exhibited more learning gain for users, but we did not use an additional questionnaire to test or validate this claim. On the contrary, Elissa and Marco [22] found that ChatGPT hindered learning, potentially due to students' inexperience with the technology.

The search system had better usability scores; the probable reason is that the click-to-wait-for-answers feature in the Chat condition is not apposite, as it blocks VDE. Search allows users to put sources to notes, which makes the insight more reliable, as also shown by Hristidis et al. [20]. Moreover, search results contained diverse content for exploration, such as charts and publications, besides texts, which may contribute to its better readability as discussed in prior work [19, 20].

However, during information retrieval, users can fail to find the answer with the search or get the wrong answer with chats. User queries in both conditions had similarities, such as iterative and drilling down to the topics, as well as differences: Besides asking for facts and reasons, queries in chat also include when- and how-type of diverse questions.

We conclude that both platforms have their merits and demerits. Users can fail information retrieval with search and retrieve unreliable information without sources using chats. We suggest combining search and chatbot (e.g., [28]) to complement each other and overcome the weaknesses so as to 1) avoid failure in information seeking, 2) enable users to retrieve the correct answer, and 3) obtain more reliable answers with sources.

5.1. Limitations

The number of participants for this pilot study is small, which hinders us from drawing firm conclusions, but the results illuminate the complementarity of the two platforms. Moreover, as the LLM tools become more user-friendly and familiar to the general public, study results can be largely affected. Follow-up studies with other data/visualization and a large general population could be conducted to expand on this investigation.

6. Conclusion

This research compares ChatGPT 4 with Vision and Google search in supporting visualization insight generation involving external evidence. We conducted a between-subjects study with 25 participants and asked them to use chat/search to complete the quantitative and qualitative insight task of the CO_2 Explorer. Results showed no significant differences between the two conditions in the task time and number/gradings of generated insights. Qualitative analysis revealed that the two systems had their own advantages and disadvantages, such as possible wrong and unreliable answers from ChatGPT and less efficient information retrieval with search. In the future, combining the two platforms will help improve both the reliability and efficiency of information retrieval.

Acknowledgments

This research is funded by the Strategic Research Council at the Research Council of Finland [Grant Number 358247].

References

[1] R. Chang, C. Ziemkiewicz, T. M. Green, W. Ribarsky, Defining insight for visual analytics, IEEE Computer Graphics and Applications 29 (2009) 14–17.

[2] P. Law, A. Endert, J. T. Stasko, Characterizing automated data insights, in: IEEE Visualization Conference - Short Papers, IEEE, 2020, pp. 171–175.

[3] R. Ding, S. Han, Y. Xu, H. Zhang, D. Zhang, QuickInsights: Quick and automatic discovery of insights from multi-dimensional data, in: the International Conference on Management of Data, ACM, 2019, pp. 317–332.

[4] P. Ma, R. Ding, S. Han, D. Zhang, MetaInsight: Automatic discovery of structured knowledge for exploratory data analysis, in: the International Conference on Management of Data, ACM, 2021, p. 1262–1274.

[5] Y. Chen, S. Barlowe, J. Yang, Click2Annotate: Automated insight externalization with rich semantics, in: IEEE Conference on Visual Analytics Science and Technology, IEEE, 2010, pp. 155–162.

[6] B. Karer, H. Hagen, D. J. Lehmann, Insight beyond numbers: The impact of qualitative factors on visual data analysis, IEEE Trans. Vis. Comput. Graph. 27 (2021) 1011–1021.

[7] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. P. Ellis, D. A. Keim, Knowledge generation model for visual analytics, IEEE Trans. Vis. Comput. Graph. 20 (2014) 1604–1613.

[8] Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung, Q. V. Do, Y. Xu, P. Fung, A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity, CoRR abs/2302.04023 (2023).

[9] W. Yang, M. Liu, Z. Wang, S. Liu, Foundation models meet visualizations: Challenges and opportunities, CoRR abs/2310.05771 (2023).

[10] OpenAI, ChatGPT 4 with Vision, https://platform.openai.com/docs/guides/vision, 2024.

[11] Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, D. Zhang, DataShot: Automatic generation of fact sheets from tabular data, IEEE Trans. Vis. Comput. Graph. 26 (2020) 895–905.

[12] Z. Cui, S. K. Badam, M. A. Yalçin, N. Elmqvist, DataSite: Proactive visual data exploration with computation of insight-based recommendations, Inf. Vis. 18 (2019).

[13] C. Liu, Y. Han, R. Jiang, X. Yuan, Advisor: Automatic visualization answer for natural-language question on tabular data, in: IEEE Pacific Visualization Symposium, IEEE, 2021, pp. 11–20.

[14] D. J. L. Lee, A. Quamar, E. Kandogan, F. Özcan, Boomerang: Proactive insight-based recommendations for guiding conversational data analysis, in: International Conference on Management of Data, ACM, 2021, pp. 2750–2754.

[15] K. Kafle, B. L. Price, S. Cohen, C. Kanan, DVQA: understanding data visualizations via question answering, in: IEEE Conference on Computer Vision and Pattern Recognition, Computer Vision Foundation / IEEE Computer Society, 2018, pp. 5648–5656.

[16] S. Latif, Z. Zhou, Y. Kim, F. Beck, N. W. Kim, Kori: Interactive synthesis of text and charts in data documents, IEEE Trans. Vis. Comput. Graph. 28 (2022) 184–194.
[17] R. Brath, C. Hagerman, Automated insights on visualizations with natural language generation, in: International Conference Information Visualisation, 2021, pp. 278–284.
[18] N. Sultanum, A. Srinivasan, DataTales: investigating the use of large language models for authoring data-driven articles, in: IEEE Visualization and Visual Analytics, IEEE, 2023, pp. 231–235.
[19] J. R. Bellinger, J. S. De La Chapa, M. W. Kwak, G. A. Ramos, D. Morrison, B. W. Kesser, BPPV Information on Google Versus AI (ChatGPT), Otolaryngology–Head and Neck Surgery (2023).
[20] V. Hristidis, N. Ruggiano, E. L. Brown, S. R. R. Ganta, S. Stewart, ChatGPT vs Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results, J Med Internet Res 25 (2023).
[21] N. F. Ayoub, Y.-J. Lee, D. Grimm, V. Divi, Head-to-Head Comparison of ChatGPT Versus Google Search for Medical Knowledge Acquisition, Otolaryngology–Head and Neck Surgery (2023).
[22] E. Arias Sosa, M. Godow, Comparing Google and ChatGPT as Assistive Tools for Students in Solving Programming Exercises (Bachelor thesis), https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-330994, 2023.
[23] M. Feng, C. Deng, E. M. Peck, L. Harrison, HindSight: Encouraging Exploration through Direct Encoding of Personal Interaction History, IEEE Trans. Vis. Comput. Graph. 23 (2017) 351–360.
[24] J. Boy, F. Détienne, J. Fekete, Storytelling in Information Visualizations: Does it Engage Users to Explore Data?, in: the CHI Conference on Human Factors in Computing Systems, ACM, 2015, pp. 1449–1458.
[25] Google, Programmable search engine, https://developers.google.com/custom-search/v1/overview, 2024.
[26] I. team, Intro.js, https://introjs.com/, 2024.
[27] J. Brooke, Sus: A quick and dirty usability scale, Usability Eval. Ind. 189 (1995).
[28] P. team, Perplexity, https://www.perplexity.ai/, 2023.

A. API Parameter settings

The settings were tested to ensure a certain amount of diversity and novelty in the chatbot's answers. The model we used is gpt-4-vision-preview with temperature 0.5, max tokens 1000, top p 1, frequency penalty 0.3, and presence penalty 0.3.

B. System Prompts

1. The initial system message: This is a visual data exploration task. The user explores CO2 emission data for [a Country Group] from 1960 to 2021, measured in tons per capita. Here is the data delimited by triple backticks in CSV format. "‘ data in CSV format "‘ The visualization

displays the data with a choropleth map, showing the geographic areas of the three countries color-coded by their CO2 emission values of a selected year, and a line chart, depicting user-selected countries' CO2 emissions from 1960 to 2021. The choropleth map uses green to red colors to code CO2 emissions from 0 to 24 tons per capita. Initially, the latest year, 2021, is selected. The line chart shows years on the x-axis and CO2 emission values on the y-axis. Initially, no country is selected."

If it is a quantitative task The user can post notes about data discoveries. The data discovery must be linked to external evidence, such as events, policies, and news. The user's task is to post as many notes about such discoveries as possible.

For a qualitative task The user can post notes. The user's task is to analyze the CO2 emission data of the three countries, coupled with the analysis of the external evidence, such as events, policies, and news, to compose a hypothesis or generalization logically and correctly as a note.

Your task is to assist the user with their task using the information provided above and your knowledge database on actual national or international news or events. Be concise with your answers.

2. The system prompts when the user selects/de-selects a country or selects a year:
   The user selects [a country name] in the line chart.
   The user de-selects [a country name] in the line chart.
   The user selects the year [year] in the choropleth map.