APPENDIX A
PROMPT USED TO GENERATE ABSTRACT UTTERANCES

We used GPT-3 (text-davinci-003) to generate abstract utterances. The prompt we used is shown in Figure 1.

```
This is a table named {{table name}}:
{{table data in CSV}}
We have a tool to automatically draw charts according to your utterances and the table data.
Now we have a clear and explicit utterance example:
{{the original utterance from nvBench}}
We want to use more abstract utterances to test the tool's ability to understand natural language, as users always give utterances that is not so explicit and clear.
Please rewrite a new abstract utterance based on the given utterance.
The new utterance should be more natural, vague, and incomplete.
Rewrite it into a
[For query] {short query consists of some keywords}.
[For question] {question ends with a question mark}.
[For command] {command that directly tells the system to do what}.
```

Fig. 1. Prompt used to generate abstract utterances. Phrasings are accommodated by selecting the For query / For question / For command options.

APPENDIX B
MODEL INPUT TEST

Due to the prompt length constraint, we can not include the complete dataset in our model input. Therefore, we mainly considered 3 kinds of information from the dataset: column name, column type, and data rows.

We tested the model performance under the different dataset information provided and summarized the result in Table I and Table II. Specifically, 'N' refers to column names, 'R' refers to data rows, and 'T' refers to column types. '2R' and '3R' refer to the top-2 data rows and top-3 data rows, respectively. We finally included column names, column types, and the top-2 data rows in our model input (N-T-2R).

TABLE I
TESTED PERFORMANCE UNDER DIFFERENT PROMPT SETTINGS.

Model Input		N	N-2R	N-3R
Top 1	BLEU	0.9140	0.9120	0.9200
	ROUGE-L	0.9084	0.9087	0.9167
	consistent	0.5899	0.6111	0.5899
Top 3	BLEU	0.9388	0.9352	**0.9458**
	ROUGE-L	0.9349	0.9315	**0.9431**
	consistent	0.7354	0.7354	0.7513
Average		0.8369	0.8390	0.8445

TABLE II
TESTED PERFORMANCE UNDER DIFFERENT PROMPT SETTINGS (CONTINUATION).

Model Input		N-T	N-T-2R	N-2R-T
Top 1	BLEU	0.9084	**0.9206**	0.9167
	ROUGE-L	0.9051	**0.9173**	0.9122
	consistent	**0.6138**	**0.6138**	0.6058
Top 3	BLEU	0.9306	0.9435	0.9372
	ROUGE-L	0.9272	0.9401	0.9337
	consistent	0.7434	**0.7672**	0.7169
Average		0.8381	**0.8504**	0.8371

APPENDIX C
MODEL INPUT ORGANIZATION

The model input organization is shown in Figure 2. Specifically, The words in '<...>' serve as a separation token. The model should output the answer sequence corresponding to step 'x' given the input. The instructions and the format of answer sequences for each step are provided in the paper.

```
{table name}
<head> {column names}
<type> {column types}
<data> {data row 1} <line> {data row 2} <line>
<utterance> {utterance}
<ans>
<sep> {Step 1 instruction} {answer 1}
...
<sep> {Step x-1 instruction} {answer x-1}
<sep> {Step x instruction}
```

Fig. 2. Model input on the step 'x'.