DATATALES: Investigating the use of Large Language Models for Authoring Data-Driven Articles

Nicole Sultanum*   Arjun Srinivasan†
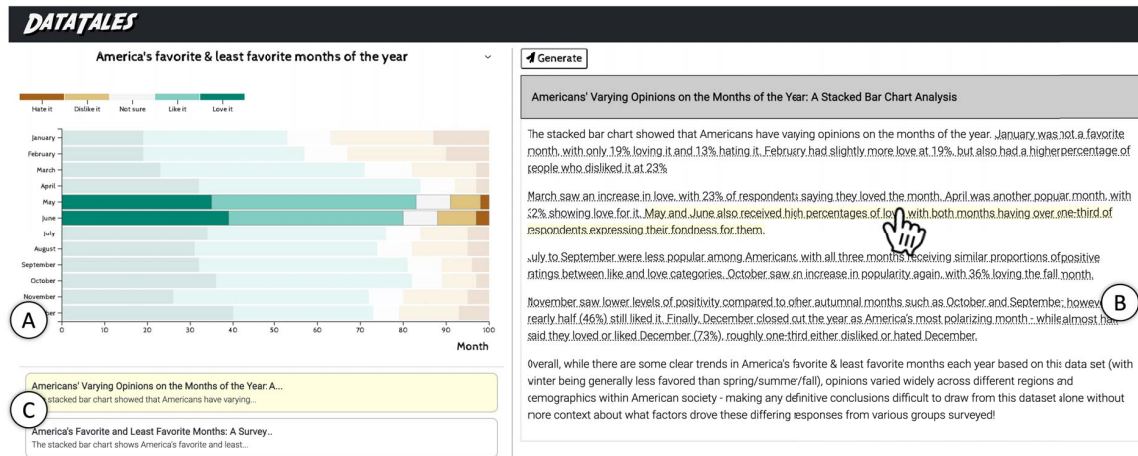
Tableau Research

Figure 1: DATATALES user interface including an interactive visualization (A), generated stories panel (B), and a history of generated stories (C). Here, as the user hovers the cursor over a sentence in the generated story, the system dynamically highlights the marks corresponding to the text (in this case, the bars for the months of May and June).

ABSTRACT

Authoring data-driven articles is a complex process requiring authors to not only analyze data for insights but also craft a cohesive narrative that effectively communicates the insights. Text generation capabilities of contemporary large language models (LLMs) present an opportunity to assist the authoring of data-driven articles and expedite the writing process. In this work, we investigate the feasibility and perceived value of leveraging LLMs to support authors of data-driven articles. We designed a prototype system, DATATALES, that leverages a LLM to generate textual narratives accompanying a given chart. Using DATATALES as a design probe, we conducted a qualitative study with 11 professionals to evaluate the concept, from which we distilled affordances and opportunities to further integrate LLMs as valuable data-driven article authoring assistants.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

Data-driven articles that feature primarily textual narratives containing claims and insights backed by data and illustrated with data visualizations are a popular means of communication in fields like journalism and business reporting [34]. Authoring data-driven articles, however, is often a complex and tedious process. Authors need to analyze the data to identify insights, order insights in an appropriate sequence, and write a cohesive narrative to communicate those insights with effective transitions and appropriate domain context.

---

*email: nsultanum@tableau.com
†email: arjunsrinivasan@tableau.com

The emergence of contemporary large language models (LLMs) and their remarkable text generation capabilities led to increased interest in assessing their value for a range of creative writing tasks [7], including data storytelling [18]. While this technology has the potential to fundamentally reshape the way people use writing tools [31], it also introduces news challenges such as unreliable outcomes, lack of domain understanding, prompt complexity, ethical concerns, among others [18]. We believe that these issues require thoughtful design solutions to circumvent them, and that different writing genres may benefit from purpose-specific features built around these models.

In this work, we investigate the potential of LLMs to support the authoring of data-driven articles. Based on the deep intertwining of charts and text in these articles, and targeting the intermediate stages of the visual storytelling process where authors are actively building a story based on exploratory findings [16], we propose *chart interaction* as a more intuitive alternative to direct prompting for conveying narrative intent to the LLM. We developed an early proof-of-concept, DATATALES, that generates textual content for an accompanying chart, and additionally allows authors add *chart annotations* to guide focus of the story. Authors can use the generated text as-is, edit portions of the text, or generate multiple instances to pick-and-choose what they like. Using DATATALES as a design probe, we conducted a qualitative study with 11 data professionals to understand the benefits and challenges of this concept to support authoring of data-driven articles. The concept was well received even in this preliminary form, showcasing potential for further investigations. We discuss lessons learned and directions for future development in the form of takeaways to share with the community.

2 RELATED WORK

Authoring data-driven articles. There is a significant body of research on narrative visualization and storytelling [16, 27, 33], a subset of which has specifically focused on authoring data-driven articles. For instance, one line of work has explored the use of
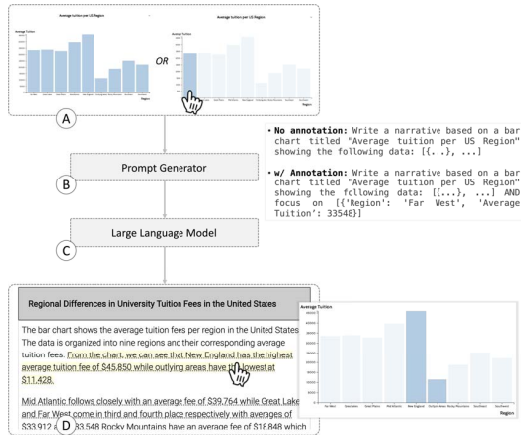
Figure 2: DATATALES workflow overview. Given a chart and an optional set of annotations, the system generates textual narratives that are interactively linked to the chart and can be further edited by authors.

markup language-based frameworks to support the authoring of interactive articles on the web [3, 13, 14]. Another set of systems like Kori [15], VizFlow [34] and DataParticles [2] adopt a more graphical and mixed-initiative approach and allow authors to configure interactive links between text and charts while authoring data articles. Besides systems that explicitly focus on content drafting and presentation fine-tuning, another body of work also includes data fact- or insight-recommendation systems that suggest singleton takeaway statements for visualizations during the data exploration phase to help authors identify talking points in their articles [19, 29, 32, 37]. Our work furthers the line of research on authoring data-driven articles by investigating the use of contemporary LLMs to generate ideas for textual content that authors can further edit.

LLMs in data visualization and writing. Recent advances in model architectures, performance, and availability have led to a surge of LLM-based applications for writing support. These include creative writing support tools such as plot suggestions [30], journalistic angle ideation [24] and co-writing of theater scripts [21]; as well as technical writing support such as argumentative writing [41], scientific writing [9], and reverse outlining for manuscript revision [4]. Within the data visualization space, LLMs have been used to power natural language interfaces [28] for visualization authoring [36]. Only a few works have looked at text content generation in a data visualization context, to create data stories from a set of user-provided keyframe data facts [35], and natural language summaries of a given chart for accessibility purposes [23]. To our knowledge, our work is the first to look specifically at leveraging LLMs for data-driven articles.

3 DATATALES

Fig. 1 shows the DATATALES user interface and Fig. 2 summarizes the system workflow. The system is implemented as a web application using a React and Python Flask setup. It features a curated list of datasets, with respective charts rendered using D3.js. For the language model, we use the OpenAI API for the 'gpt-3.5-turbo' model [1]. Below we detail workflow steps and core system features.

Chart and user annotations. DATATALES covers a wide array of charts commonly found in data-driven reports and articles, including bar charts with variants like stacked and group bars, scatterplots, single- and multi-series line charts, and choropleth maps. The current implementation contains a set of predefined charts covering a breadth of datasets including demographic survey responses, unemployment rates, automobile data, and Olympic medal winner

---

[1]The latest model available for development at the time of this research.

history, among others. When asking the system to generate a story (via the `Generate` button, Fig. 1B), authors can have the entire chart considered for input or optionally add annotations to guide the LLM to emphasize specific data points or ranges when generating its response. For instance, Fig. 2A shows an example where an author highlights two bars they want the system to focus on when generating a story. DATATALES supports various annotations including mark selection, color legend range selection, and axis range selection, which can be combined for more complex guidance [25].

Prompt generation. The key idea underpinning DATATALES is that a system can take a chart or an annotated chart as input and leverage a LLM to recommend data-driven narratives. To this end, we iterated on several template variations to generate the prompts that are fed into the LLM. Specifically, we explored different features to include (or exclude) in the prompt such as the chart type, encodings, analytic tasks associated with specific charts [10, 26, 39], the chart title, the underlying dataset metadata, and user annotations, story length, among others. Besides experimenting with features, we also tried different phrasings to assess if the order of features or the grammar of the prompt notably impacted the generated narrative.

We generated 10-20 narratives for each chart type with different combinations of these features. Inspecting the results, we iteratively excluded or combined features that yielded redundant results. For instance, we noticed that including the encoding information in the prompt generated statements reiterating the chart. We thus excluded encoding details from the prompt as such statements tend to offer little value to readers [20]. Similarly, we initially experimented with including analytic tasks (e.g., finding extremes, identifying correlations). However, we noticed that including the chart type in the prompt (e.g., 'bar chart', 'scatterplot', 'line chart') resulted in narratives comparable to those generated by including analytic tasks. Correspondingly, keeping in mind the simplicity and brevity of specifying the chart type (over analytic tasks), we only included that information in the final prompt template. Fig. 2B shows an example of the prompts generated by DATATALES but the general template for generating data narratives is as follows:

```
Write a narrative based on a [chartType] showing
the following data: [chartData] on the topic
"[chartTitle]" focusing on:  [chartAnnotations*]
```

where, `*` indicates an optional parameter that is included in the prompt only if it is available in the input chart. `chartData` is the data array that is bound to the marks and `chartAnnotations` is a list of data items for selection annotations (e.g., $\{Year : 2000, Country : Australia\}$) and/or values in the case of axis brush annotations (e.g., $\{Year\ between\ [1980, 2001]\}$). Once a narrative is generated, we prompt the LLM again to generate a title:

```
Suggest a title for the following narrative:
[narrativeText].
```

The title and text are sent as to the system front-end as a self-contained story. These prompts generated reasonable results for our purposes, although we argue that further experimentation with prompt patterns [38] would be worthwhile.

Linking the generated text to the input chart. Once the LLM generates the narrative, DATATALES proactively processes the generated story to identify data references. Similar to prior natural language systems for visualization (e.g., [8, 22]), we use a combination of dependency parsing and keyword matching to map phrases in a sentence to attributes and values in the visualized data. DATATALES highlights whole sentences containing data references using a dotted underline to emphasize that the sentence talks about a specific set of marks on the chart. To aid reading and comprehension, and incorporating ideas from prior work on interactively linking text and charts [12, 15, 32, 34], as authors hover on these underlined sentences, DATATALES highlights relevant portions of the chart (see
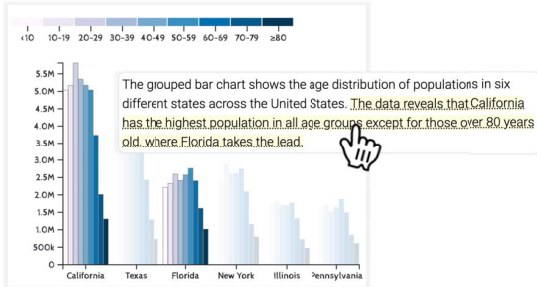
Figure 3: Example of an incorrect statement generated by the LLM (contrary to the text, the chart shows that Florida *does not* have a higher number of people over the age of 80 compared to California). The text→chart linking feature helps verify the statement and identify the erroneous interpretation by dynamically highlighting the two states.

Figs. 1 and 2D). Besides improving readability, our motivation to include this text→chart linking was also that visually seeing the data being referred to in the text could serve as a quick verification for potential hallucinations or incorrect interpretations by the LLM (e.g., Fig. 3). Authors can then redact the stories themselves, and their edits are shown in a different italicized format.

4 EVALUATION

To assess whether our envisioned concept of using chart interaction for LLM story generation made sense in the context of data-driven articles, we conducted a qualitative user study. We used DATATALES as a *design probe* to expose participants to this concept in the context of an authoring task, and then elicited feedback on the experience.

We recruited 11 data professionals (P1-P11) with prior experience in authoring data-driven articles or similar reports. Backgrounds encompassed content writers, dashboard designers, project managers and consultants, spanning multiple organizations. Participants were recruited via slack communities on related interest channels. Incidentally, most participants had some prior exposure to LLM-based tools, and five reported using these tools for authoring support at some point (e.g., for brainstorming, starting points, outlines, summaries).

Feedback sessions entailed a 20-min data story authoring task on a given chart followed by a semi-structured interview to discuss their experiences with the tool. In lieu of conducting data analysis from scratch, participants were given one of a subset of four distinct dataset+chart types available in the tool (including stacked bar chart, line chart, scatterplot, and choropleth), which helped standardize experiences and keep sessions concise. Participants were told to use DATATALES in their authoring process however they saw fit, while editing their working draft in a separate document editor for maximum editing flexibility. We also encouraged them to think aloud whenever possible during the authoring task.

This study setup provided us with rich qualitative data to assess potential and limitations of LLMs for this task. We organize our findings in the form of **takeaways (T1-T13)**, encompassing observations of authoring workflows (Section 4.1), perceived value and affordances of LLM-based article authoring systems like DATATALES for data-driven authoring experiences (Section 4.2), and identified limitations plus potential solutions (Section 4.3).

4.1 Authoring Workflows

Lessons and ideas emerging from task observations are as follows.

(T1) A master draft + multiple stories. Participants were free to decide a format and framing for their stories, which led to a diverse set of data-driven articles. That said, authoring workflows were fairly consistent across participants: they all generated multiple versions, reused chunks from one or more versions — occasionally a whole story, but most frequently short paragraphs from various ones — rearranged them, and then made editorial revisions for style and

flow. This suggests a setup for managing multiple story generation outcomes, and that maintaining an integrated master draft which can be easily populated with generated segments would be useful.

(T2) Expediting error checks. As expected of LLM output, a number of inaccuracies [1] were spotted, prompting them to carefully check generated text for errors: *"is this legit?"* (P3). The text-chart highlights were frequently used for this purpose and several agreed on its usefulness (P3, P8, P9), suggesting text-chart readability aids [15] should be further explored. On that note, some folks appreciated how author changes were explicitly signaled (P2, P5), helping retain context of what was fixed and what needs checking.

(T3) Synergies between chart and text. Participants extensively leveraged chart interaction for their story generation: from an average of 4 stories per person, about 3 featured annotations. While the annotated stories did not always feature the depth and framing participants were hoping for (more under **T9**, **T10**), it consistently matched the selections (and author intent, as per think-aloud feedback), and results were still often usable and repurposed. Some participants also explicitly acknowledged the value of having the chart integrated into their drafting environment (P6, P7). These findings suggest that DATATALES's use of chart interaction for text generation shows promise and is worth exploring further.

(T4) Coupling of annotations and generated stories. Annotations were frequently used to get more details on selected data features, usage that was largely intuitive and represented important context for the text. We posit that preserving annotation context for generated snippets on a master draft would be very useful, not only for authors to recall provenance of snippets but also to potentially reuse as embedded highlights for readers, e.g., in the context of a dynamic story format such as scrollytelling [34].

(T5) Potentially time saving. While participants were not expected to finish their stories within the 20min, 3 of them successfully completed a first draft in the allotted time, suggesting potential efficiency gains in the authoring process. Several participants could also foresee saving time in the long run, e.g., *"would cut out a good 15 to 20 minutes of my work"* (P6), and getting a head start on the writing, e.g., *"Getting started is sometimes hardest thing (...). I'll be looking at the data, procrastinating, trying to find correlations and relationships(...). And it does that for me, at least a base level"* (P8).

4.2 Affordances

Despite the limited nature of the tool as a proof-of-concept design probe, participant reactions to the experience ranged from congenial to enthusiastic. Rationales on how DATATALES supported their authoring experience in new and positive ways are compiled below.

(T6) Insights over data facts. While data facts are an important part of a data story, the segments most often repurposed and appreciated by participants were those containing level-3 and level-4 statements in Lundgard and Satyanarayan's categorization of chart descriptions [20], which participants referred to as "*the why's*" (P3, P10, P11). For example, on a dataset about cars acceleration vs. horsepower vs. country of origin, this could include things like identifying trends (e.g., *"cars with higher horsepower tend to have better acceleration rates"*), conclusions following findings (e.g., *"The US auto market prioritizes higher horsepower"*), and external context (e.g., *"policymakers should consider regulating emissions for consumers who value speed over efficiency"*). Several added that aggregating this *"human knowledge"* was one of the most valuable aspects of the experience, complementing their authoring work with new information (P1, P3), alternative framings (P3, P7), and confirmation of current viewpoints (P3, P11).

(T7) Explanatory support: what to talk about and how. Getting a first initial draft is challenging, and having a range of full stories available helped provide starting points to overcoming writer's block (P6, P8, P10, P11): *"I didn't know where I wanted to approach it, and then after generating a couple stories, I saw a trend and de-*

cided that'd be my focus" (P6). New ideas or context present in those stories also provided inspiration for new directions to explore (P5, P7), e.g., the extent that *"the Great Depression"* affected US gold medal performance in the Summer Olympics (P7); as well as seeing familiar findings or terms portrayed in a different ways, with evocative phrasings, e.g., *"powerful muscle cars"* (P11), and unique ordering of findings (P3). We argue there may be value in not only supporting easy generation and management of different stories, but also allowing for more diversity across different stories (e.g., via a slider to control the model's `temperature`), even at the cost of more spurious findings.

(T8) Exploratory support: a different lens on the data. An unexpected use of DATATALES was as a data exploration tool, e.g., to form hypothesis (P3), to gather facts (P7) and to get a high-level summary of the data in natural language form (P2, P6, P9). While our study setup induced some analysis as participants were asked to work with an unknown dataset, several of the dynamics observed applied to in-between stages of analysis and storytelling: e.g., getting ideas for additional datasets and facts to look into (P7, P8), and using annotations to dig deeper into individual data points (P11). Several participants also leveraged DATATALES to "test" hypotheses by confirming or denying prior assumptions (P3, P7, P10), and to *"ask its opinion"* (P7, P11). This showcases the intertwined nature of analysis and storytelling, and how authoring tasks can benefit from interactive visualizations integrated into the drafting environment.

4.3 Opportunities

While overall reactions to the tool were net-positive, participants also raised several concerns and suggestions for improvement, informing many compelling directions for future work.

(T9) More control over overall style. A prominent pain point was the lack of control over *voice* (e.g., corporate voice (P5), a business owner's perspective (P3)), *tone* (e.g., formal vs. personal (P2), make it less "robotic" (P8)), and *format* (e.g., organize findings from highest to lowest counts (P4)). Generated stories were often found *"too wordy"* (P5-P7, P10, P11), requiring heavy editing to cut them down. Those with prior LLM exposure suggested bridging this gap by writing or editing underlying prompts generated by the tool (P5, P6, P10); on the other hand, it was remarked that prompting could be found too intimidating or unfamiliar for other authors (P3), which calls for some form of reasonable middle ground. We envision DATATALES could provide predefined fields for authors to describe target audience, voice, and format in natural language, which could then be incorporated into the base template to generate a new story.

(T10) Co-writing micro-tasks. Apart from overall style, participants also wanted assistance in generating paragraph-level content for targeted insights (P4, P5, P6), e.g., *"let's focus on January thru March, and talk about the holiday angle"* (P5) and *"tell me why the US has such an outlier number of cars on the higher horsepower end"* (P6). While power-users could again benefit from direct prompting here, P5 suggested a more seamless *fill-in-the-blanks* approach, designating spaces for automatic completion for diegetic prompting [5], i.e., leveraging parts of the writing itself. For data-centered insights, we envision combining chart interaction with natural language input, e.g., by editing the chart title to nudge generation of different insights, drawing trend lines over the chart to indicate what patterns are important to focus on, and adding descriptive annotations to particular data points to retrieve targeted context.

(T11) More exploratory aids. Many participants wanted to take their insights a step further by getting recommendations for external related datasets to include in the story (P3-P7, P11), and potential comparisons to be drawn within or across datasets (P2, P4-P7, P11). Some also framed these recommendations as an opportunity to continue learning about the data **(T8)** and helping them formulate *"follow up questions"* (P3) and *"additional areas of exploration"* (P5), giving authors more directions to consider for their stories.

(T12) Summarize stories. Another consequence of "wordy" stories **(T9)** is they take *"a lot of time to read through"* (P5), which is specially onerous when trying to generate stories for exploratory analysis purposes **(T8)**, i.e., natural language overviews of the data. Some participants suggested an option to depict stories in a more concise format (P2, P5, P7, P11), e.g., bullet-point form (P2, P5). Beyond data analysis overview, these summaries could also allow for quicker inspection of new stories, faster recall of past stories, and support revisions of a master draft via reverse outlining [4, 11].

(T13) Internal knowledge, external validation. As stated earlier, a number of inaccuracies were surfaced in the generated stories **(T2)**. These include objective errors (e.g., such as hallucinations, failure to spot obvious patterns, labeling mix-ups), but also subtle mistakes that are hard to spot without expert background: e.g., on a misguided reasoning for Vermont's low unemployment rates, P8 warned *"if I didn't have local (Vermont) knowledge, I'd just have taken this at face value"*. To mitigate this, automatic retrieval of supporting references was suggested, e.g., source citations (P9) and external datasets for triangulation (P8, P9). Some participants also wanted to provide corrective feedback to the model (P3, P6), on the assumption that it would get better *"the more you use it"* (P6). Another related concern was trying to sort out what statements are derived from the underlying chart data and what was brought in by the LLM as external context (P2, P5, P9). Similar to how manual edits are clearly marked **(T2)**, participants suggested further signaling passages with a different color or formatting to denote source.

5 REFLECTIONS

Following our takeaways **(T1-T13)**, we conclude with a few thoughts that emerged from discussions that are pertinent to the use of LLMs for authoring and visual storytelling more generally.

First, we consider **bias versus framing.** One of the challenges reported by participants was how difficult it was at times to get generated text that matched the points they wanted to make in a story, i.e., *"confirm their viewpoints"*. On one hand, *"picking a side"* is an integral exercise of telling a story. On the other, we also know how prone LLMs are to fabricating evidence in a convincing manner. As such, to honor the overall purpose of data stories to stay true to the data (within the many valid interpretations that may arise), and given the significant exploratory component that the authoring process entailed, it is essential we consider ways to mitigate confirmation bias and amplification of harmful speech.

A related point is the importance of **bridging data literacy**. It was raised in participant discussions how folks with lower data literacy skills would benefit the most from authoring support (P5, P9), but at the same time, having these skills would be crucial to properly guard against LLM errors (P8). Under the same tenet of mitigating misinformation, we should consider ways to better equip authors with critical thinking skills: from facilitating fact checking and encouraging all statements to be fact-checked, to nudging techniques that encourage reflection, e.g., phrasings that "ask a question" instead of stating "what and why" [6].

And finally, we make a case for the use of **LLMs alongside classic techniques.** While our studies demonstrated how LLMs can add value to the authoring by easily incorporating context and insights, their unpredictability does offset some of the efficiency it is purported to bring. Many of the solutions suggested to deal with this uncertainty would call for add-on modules to parse the generated text, identify entities, link text entities and chart elements, retrieve external evidence, and so on. While there is ongoing work in applying LLMs to natural language tasks like information extraction [17] and fact-checking [40], we argue that tried and tested heuristics-based algorithms and classic NLP techniques will still play an essential role in the foreseeable future.

REFERENCES

[1] A. Borji. A categorical archive of chatgpt failures. *arXiv preprint arXiv:2302.03494*, 2023.

[2] Y. Cao, J. L. E, Z. Chen, and H. Xia. Dataparticles: Block-based and language-oriented authoring of animated unit visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2023.

[3] M. Conlen, M. Vo, A. Tan, and J. Heer. Idyll studio: A structured editor for authoring interactive & data-driven articles. In *ACM Symposium on User Interface Software and Technology*, pp. 1–12, 2021.

[4] H. Dang, K. Benharrak, F. Lehmann, and D. Buschek. Beyond text generation: Supporting writers with continuous automatic text summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–13, 2022.

[5] H. Dang, S. Goller, F. Lehmann, and D. Buschek. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2023.

[6] V. Danry, P. Pataranutaporn, Y. Mao, and P. Maes. Don't just tell me, ask me: Ai systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal ai explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2023.

[7] X. Fang, D. T. K. Ng, J. K. L. Leung, and S. K. W. Chu. A systematic review of artificial intelligence technologies used for story writing. *Education and Information Technologies*, pp. 1–37, 2023.

[8] T. Gao, M. Dontcheva, E. Adar, Z. Liu, and K. G. Karahalios. Data-tone: Managing ambiguity in natural language interfaces for data visualization. In *Proceedings of the 28th annual acm symposium on user interface software & technology*, pp. 489–500, 2015.

[9] K. I. Gero, V. Liu, and L. Chilton. Sparks: Inspiration for science writing using language models. In *Designing Interactive Systems Conference*, pp. 1002–1019, 2022.

[10] Y. Kim and J. Heer. Assessing effects of task and data distribution on the effectiveness of visual encodings. In *Computer Graphics Forum*, vol. 37, pp. 157–167. Wiley Online Library, 2018.

[11] C. L. King. Reverse outlining: A method for effective revision of document structure. *IEEE Transactions on Professional Communication*, 55(3):254–261, 2012.

[12] N. Kong, M. A. Hearst, and M. Agrawala. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 31–40, 2014.

[13] S. Latif, D. Liu, and F. Beck. Exploring interactive linking between text and visualization. In *EuroVis (Short Papers)*, 2018.

[14] S. Latif, K. Su, and F. Beck. Authoring combined textual and visual descriptions of graph data. In *EuroVis (Short Papers)*, 2019.

[15] S. Latif, Z. Zhou, Y. Kim, F. Beck, and N. W. Kim. Kori: Interactive synthesis of text and charts in data documents. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):184–194, 2021.

[16] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*, 35(5):84–90, 2015.

[17] B. Li, G. Fang, Y. Yang, Q. Wang, W. Ye, W. Zhao, and S. Zhang. Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness. *arXiv preprint arXiv:2304.11633*, 2023.

[18] H. Li, Y. Wang, Q. V. Liao, and H. Qu. Why is ai not a panacea for data workers? an interview study on human-ai collaboration in data storytelling. *arXiv preprint arXiv:2304.08366*, 2023.

[19] H. Li, L. Ying, H. Zhang, Y. Wu, H. Qu, and Y. Wang. Notable: On-the-fly assistant for data storytelling in computational notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023.

[20] A. Lundgard and A. Satyanarayan. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1073–1083, 2021.

[21] P. Mirowski, K. W. Mathewson, J. Pittman, and R. Evans. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–34, 2023.

[22] A. Narechania, A. Srinivasan, and J. Stasko. Nl4dv: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):369–379, 2020.

[23] J. Obeid and E. Hoque. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model. *arXiv preprint arXiv:2010.09142*, 2020.

[24] S. Petridis, N. Diakopoulos, K. Crowston, M. Hansen, K. Henderson, S. Jastrzebski, J. V. Nickerson, and L. B. Chilton. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023.

[25] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 230–239. IEEE, 2017.

[26] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2505–2512, 2018.

[27] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[28] L. Shen, E. Shen, Y. Luo, X. Yang, X. Hu, X. Zhang, Z. Tai, and J. Wang. Towards natural language interfaces for data visualization: A survey. *arXiv preprint arXiv:2109.03506*, 2021.

[29] D. Shi, X. Xu, F. Sun, Y. Shi, and N. Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2020.

[30] N. Singh, G. Bernal, D. Savchenko, and E. Glassman. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction*, 2022.

[31] J. Spencer. The future of writing in a world of artificial intelligence. https://spencerauthor.com/ai-essay/, 2023.

[32] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):672–681, 2018.

[33] C. D. Stolper, B. Lee, N. H. Riche, and J. Stasko. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. 2016.

[34] N. Sultanum, F. Chevalier, Z. Bylinskii, and Z. Liu. Leveraging text-chart links to support authoring of data-driven articles with vizflow. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2021.

[35] M. Sun, L. Cai, W. Cui, Y. Wu, Y. Shi, and N. Cao. Erato: Cooperative data story editing via fact interpolation. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):983–993, 2022.

[36] Y. Wang, Z. Hou, L. Shen, T. Wu, J. Wang, H. Huang, H. Zhang, and D. Zhang. Towards natural language-based visualization authoring. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1222–1232, 2022.

[37] Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, and D. Zhang. Datashot: Automatic generation of fact sheets from tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):895–905, 2019.

[38] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt. A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*, 2023.

[39] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & cognition*, 27:1073–1079, 1999.

[40] T. Zhang, H. Luo, Y.-S. Chuang, W. Fang, L. Gaitskell, T. Hartvigsen, X. Wu, D. Fox, H. Meng, and J. Glass. Interpretable unified language checking. *arXiv preprint arXiv:2304.03728*, 2023.

[41] Z. Zhang, J. Gao, R. S. Dhaliwal, and T. J.-J. Li. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. *arXiv preprint arXiv:2304.07810*, 2023.