

Composing Data Stories with Meta Relations

Haotian Li, Lu Ying, Leixian Shen, Yun Wang, Yingcai Wu, and Huamin Qu

Abstract—To facilitate the creation of compelling and engaging data stories, AI-powered tools have been introduced to automate the three stages in the workflow: analyzing data, organizing findings, and creating visuals. However, these tools rely on data-level information to derive inflexible relations between findings. Therefore, they often create one-size-fits-all data stories. Differently, our formative study reveals that humans heavily rely on meta relations between these findings from diverse domain knowledge and narrative intent, going beyond datasets, to compose their findings into stylized data stories. Such a gap indicates the importance of introducing meta relations to elevate AI-created stories to a satisfactory level. Though necessary, it is still unclear where and how AI should be involved in working with humans on meta relations. To answer the question, we conducted an exploratory user study with Remex, an AI-powered data storytelling tool that suggests meta relations in the analysis stage and applies meta relations for data story organization. The user study reveals various findings about introducing AI for meta relations into the storytelling workflow, such as the benefit of considering meta relations and their diverse expected usage scenarios. Finally, the paper concludes with lessons and suggestions about applying meta relations to compose data stories to hopefully inspire future research.

Index Terms—Visualization, Data Storytelling, Human-AI Collaboration

I. INTRODUCTION

DATA stories appear everywhere from public communication [2], business presentations [3], to personal reflection [4]. They present multiple *story pieces* supported by data with coherent *relations* between them to convey messages between authors and creators [5], [6]. To facilitate data story creation, introducing artificial intelligence (AI) assistance has gained interest from both researchers and practitioners [7], [8].

When creating data stories, it is crucial to connect multiple data story pieces to convey a complete and clear message [9]. However, how to connect story pieces receives little attention in existing AI-powered data storytelling tools. Following the seminal research in story piece organization [10], [11], most existing methods rely on data relations among story pieces to assist human creators. For example, DataShot [12] groups data facts with similar data attributes or filters together and generates data posters accordingly. Calliope [13] infers the relations, such as contrast and similarity, between story pieces based on their data attributes. Then, it searches for the best sequence of these story pieces based on their relationship. A

H. Li, L. Shen, and H. Qu are with The Hong Kong University of Science and Technology. E-mail: {haotian.li, lshenaj}@connect.ust.hk; huamin@cse.ust.hk

L. Ying and Y. Wu are with the State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China. E-mail: {yingluu, ycwu}@zju.edu.cn

Y. Wang is with Microsoft. E-mail: wangyun@microsoft.com.

This paper is part of the thesis “Bridging Data Analysis And Storytelling With Human-AI Collaborative Tools” by the first author H. Li [1].

more recent research, Socrates [14] relies on the data relations to organize story pieces with consideration of users’ feedback.

To create coherent and clear data stories, humans often consider more factors beyond facts and relations from data, such as the social background, the goal of data stories, and also audiences’ preference [9], [15], [16], which are overlooked by existing AI-powered tools. As a result, the stories created by AI-powered tools lack the consideration of humans’ preferences and expectations and are homogeneous when the data facts are similar. Human creators still need to spend considerable effort in customizing them according to their requirements. To fill the gap, we introduce *meta relations* to reflect the connection between story pieces more comprehensively. Different from data relations that can be inferred from story pieces solely, meta relations describe the relations that are from the meta information of data stories, such as background information of these stories and analytical questions that the data story aims to answer. Incorporating such relations can enable AI to understand human creators’ mindsets and, therefore, organize data stories closer to their expectations.

An example of considering meta relations in data storytelling is from an award-winning data article “Scientific Proof that Americans are Completely Addicted to Trucks” from Bloomberg, where the auto sales in 2014 are introduced [17]. The article introduces the fact of the decreasing sales of the hybrid electric car model, Toyota Prius, and the facts of the increasing sales of other plug-in and pure electric car models, including Nissan Leaf, Tesla S, etc., consecutively. Since meta relations about these car models, *i.e.*, *they belong to different electric car categories with competitive natures*, are considered, the article compares these car models to convey the insightful message of a growing interest in plug-in and pure electric cars of the market. However, if considering the facts solely or referring to the dataset of the article shown at the bottom, we notice that these cars do not share any commonalities in categorical features, indicating a weak data relation among them. When only such relations are considered, these facts are unlikely to be introduced together. The example clearly demonstrates that meta relations between data facts are indispensable to a well-organized data story. Such findings motivate us to explore how to introduce meta relations to AI-powered data storytelling tools to compose stories that align with humans’ knowledge and preferences.

In our research, we first conducted a formative study to enhance our understanding of meta relations in practice. From these responses, we summarized findings about meta relations, including the categorization of diverse meta relations based on the source types (*i.e.*, domain knowledge and narrative intent) and their challenges with meta relations. These lessons further informed us of the considerations in developing an AI-powered

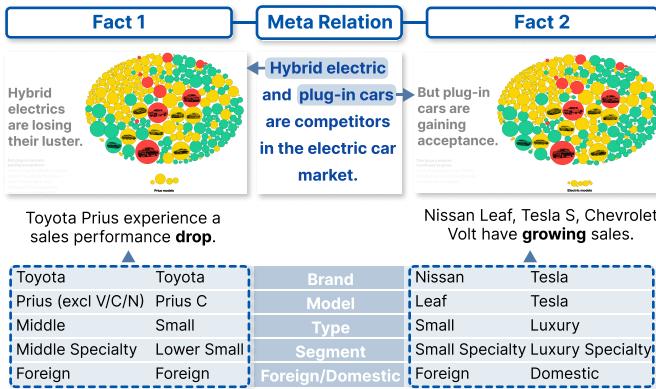


Fig. 1. A motivating example to showcase the necessity of meta relations in data storytelling from Bloomberg [17].

tool for composing data stories with meta relations, such as allowing free-form meta relation definition and assisting users in identifying meta relations. However, it is still unclear where and how humans would like to receive assistance from AI on meta relations in data story creation workflow. To figure out the issue, we developed a prototype tool, Remex, and conducted an exploratory user study with it. Remex enables inferring meta relations between story pieces automatically for users' selection and then organizing story pieces into data stories with joint consideration of meta and data relations automatically. The participants in the user study were instructed to create data stories with both Remex and Notable [18], another notebook tool that supports data storytelling based on data relations alone, without considering meta relations. According to participants' feedback, we summarized a series of findings about applying meta relations in data storytelling. First, they confirmed the usefulness of introducing meta relations into the storytelling workflow explicitly, such as providing hints about the direction of data stories and connecting story pieces more tightly. Second, we summarized various desired usage of meta relations, like showing underlying relations between two selected facts and providing suggestions for next story pieces. Furthermore, we also noticed challenges in delivering meta relations, such as displaying meta relations with minimized mental load. We also provide lessons learned in the research and suggestions for future AI-powered tools that aim to incorporate meta relations.

To conclude, our contributions include:

- **introducing meta relations** to enhance the connections between story pieces and facilitate automatic data storytelling.
- **unveiling meta relations and pain points associated with them** through a formative study with data workers.
- **summarizing design considerations** for tools that integrate meta relations into the storytelling workflow and exemplify the considerations with a proof-of-concept tool, Remex.
- **outlining the benefits and usage scenarios of meta relations** based on a user study where participants have hands-on experiences to compare Remex with a baseline without meta relations.

- **summarizing the lessons** for researchers to enhance the consideration of meta relations in future studies.

II. RELATED WORK

We review the prior literature from three perspectives: AI-powered data storytelling tools (Sec. II-A), automatic data story piece organization (Sec. II-B), and meta information in data storytelling (Sec. II-C).

A. AI-powered Data Storytelling

Telling coherent and compelling data stories often requires diverse skills, such as visualizing data and drafting scripts, and considerable human efforts due to its long pipeline [5], [19]. To democratize data storytelling, researchers and practitioners have explored how to leverage the power of AI techniques in the whole workflow [8], [9].

Recent research has proposed to characterize existing AI-powered data storytelling tools according to the stages where the tools facilitate, including analysis, planning, implementation, and communication [9]. The tools enhancing the *analysis* stage mainly mine data facts from datasets to assist human creators or provide materials for story generation. For example, Shi *et al.* [13] and Lu *et al.* [20] generate data stories in the scrollytelling form based on mined data facts in the analysis stage. Erato [21] and Notable [18] suggest potential data facts for humans' selection. In the *planning* stage, the data story creators need to consider how to arrange the story pieces about data findings into a coherent story. Representative tools that cover this stage include DataShot [12] and Calliope [13]. Sec. II-B provides a more comprehensive introduction to how these tools organize data story pieces. The *implementation* stage has received the most attention from researchers. The tools often assisted in creating static or dynamic visualizations. For example, DataQuilt [22] extracts reusable visual elements with AI techniques to facilitate new infographic creation. Text-to-Viz [23] generates infographics according to users' natural language descriptions. More recently, Ying *et al.* [24] and Shen *et al.* [25] proposed approaches to generate data videos automatically from static charts with assistance from LLMs. Only a few tools are proposed to enhance the *communication* phase where data stories are directly communicated with the audiences. Lee *et al.* [26] and Hall *et al.* [27] facilitate more vivid communication of data stories through recognizing storytellers' gestures and presenting corresponding story pieces automatically.

Following the AI-powered data storytelling trend, to enable AI to better understand the intentions of human creators, this paper goes beyond data-level information and explores the integration of meta relations (e.g., domain knowledge and story's narrative intents) beyond datasets. As part of our research, we developed a proof-of-concept tool called Remex, which incorporates meta relations into the workflow of communicating data findings in notebooks with slides, covering analysis, planning, and implementation stages and emphasizes how to organize data stories with meta relations in the planning stage more conveniently and effectively. In the next section, we will introduce how previous tools achieve

automatic data story organization and the aspect that meta relations would like to improve.

B. Automatic Data Story Piece Organization

When authoring data stories with multiple data facts, it is necessary to organize these data facts coherently [9]. Previous research has identified various perspectives to be considered when organizing data stories, such as narrative strategies [28], audiences' perceived mental load [10], [11], and the meta information of the story [7], [15]. It is demanding to consider these factors comprehensively. To address the challenge, automatic approaches have been proposed.

As pioneer research in automating data story organization, Hullman *et al.* [10] identified a set of common connection patterns between adjacent charts in data stories, mostly from the data perspective, such as dimension and measure changes. Then they proposed a graph-based approach to arrange charts based on the relations between their data attributes to minimize audiences' transition costs. Kim *et al.* [11] proposed Graph-Scape to enhance Hullman *et al.* by enriching the data relations and developing a workable model based on Vega-Lite [29]. Following their approaches, later research still relies on data relations to organize story pieces automatically. DataShot [12] organizes data facts into groups based on their shared topics, such as common filters. Calliope [13] proposes a group of data relations between data facts, including contrast and similarity, and organizes data stories based on such relations with a Monte Carlo Tree Search. A similar approach is adopted by Socrates [14] with an additional consideration of humans' feedback. Erato [21] augments Calliope by allowing human creators to fix an important fact sequence first and then fill in additional facts to enhance humans' agency in automatic data story organization.

Comparing the factors considered by existing approaches and humans' practices, we conclude that the previously proposed approaches have a common drawback where they consider merely objective and inflexible data relations. They can result in "one-size-fits-all" data stories and fail to cater to users' specific needs according to their backgrounds or preferences, as indicated in previous research [9], [18]. To take one step further, we explore how to incorporate more flexible and customized meta relations from metadata of datasets to provide effective assistance to story piece organization automatically.

C. Meta Information in Data Storytelling

Our paper introduces meta relations as a new perspective to delineate the connection between data facts. They are extracted from meta information of data stories, *i.e.*, domain knowledge and narrative intent. The introduction of meta relations is motivated by previous research about meta information in data storytelling.

According to Chevalier *et al.* [19], metadata was necessary for data journalists to understand the data before making stories. Hullman and Diakopoulos [30] argued that the narrative intent of visualization creators affected the final data story. Li *et al.* [7] indicated that the meta information of datasets,

such as social context and project background, was often considered when making data stories. Furthermore, AI assistance for meta information collection in data storytelling is desired. Besides the importance when making stories, meta information can also affect data story comprehension. Burns *et al.* [15] found that the metadata of communicative visualizations (*e.g.*, dataset descriptions and the goal of visualizations) could enhance the perceived thoroughness of readers and help them assess whether the visualization is trustworthy. Lin *et al.* [16] mentioned that domain knowledge of datasets could affect the understanding of data insights.

These studies inform that meta information is crucial for both creating and understanding data stories, though it was seldom considered in previous research on AI-powered data storytelling. In our paper, we propose meta relations as a potential way to enhance the consideration of meta information in data story creation. Our study aims to understand the practices and expected AI assistance for meta relations.

III. FORMATIVE STUDY

To learn about the practices, pain points, and expectations of meta relations, we conducted a formative study with six data workers. In our paper, data workers are those who conduct data science work in daily jobs regardless of whether they are professionals or amateurs [31], [32].

A. Participants

We invited six participants (4 males, 2 females, $Age_{mean} = 26.33$, $Age_{std} = 0.75$, denoted as P1-P6) from various disciplines, including computer science (P1, P2, P5), information design (P3), economics (P4), and chemistry (P6) through our professional network. We recruited participants with diverse backgrounds to enhance our understanding of meta relations with multiple viewpoints from various data workers. They have rich experience in analyzing and communicating data ($Experience_{mean} = 5.17$ years, $Experience_{std} = 2.29$ years).

B. Procedure

Before the study, we first briefly introduced the purpose of the research. After collecting the consent of recording and using participants' demographic information and opinions expressed in the study, we first enquired about the participants' background information, such as their ages, genders, and work nature. Then we started our study with a brief introduction to data storytelling, where we explained the terms such as story pieces and meta relations. Next, the participants were asked to share their practices in identifying and applying meta relations when communicating data findings. To help them recall the experiences, we also designed a task where the participants were instructed to compile a data story to introduce SUV car sales. Specifically, we provided twelve data facts about SUV cars from the widely applied car sales dataset in previous data storytelling research [12], [18]. The dataset includes five attributes: car models, brands, categories, sales, and years. We also provided background information about

the SUV models from Wikipedia. They were instructed to identify meta relations between 12 data facts and organize them into a satisfactory data story using around 15-20 minutes. During the study, they were encouraged to express how they decided on the story organization. After finishing the task, we asked them to introduce the story and meta relations inside. Finally, we asked the participants about their challenges with meta relations and what assistance they expect to enhance telling data stories with meta relations. The study with one participant lasted for around 40 minutes. The whole study was recorded and transcribed. Following the analysis approach in previous research [18], after interviews, the first author organized the participants' opinions based on the notes and recordings. Then, the results were discussed with co-authors to reach a consensus. Finally, we identified two key findings from the study as below.

C. Results

Based on participants' opinions, we summarized two key results about the practices regarding meta relations.

The types of meta relations are diverse and are mostly identified based on their sources. Overall, all participants confirmed that they often took meta relations into consideration when performing analysis and follow-up communication. When introducing the meta relations they often consider, the participants often categorize them based on the sources of meta relations. We notice that there are two types of sources from which meta relations come, including *domain knowledge* (P1, P2, P3, P4, P5, P6) and *narrative intent* (P3, P5).

Domain knowledge refers to the expertise related to a specific domain that is not included in the dataset, *e.g.*, the origins of cars in the car sales dataset. Domain knowledge is the most commonly mentioned type of meta relation source. For example, P6 organized the facts in our task using the size of SUV cars (*e.g.*, mid-size and full-size) based on his knowledge. Similarly, P5 considered BMW models to be representative of luxury models and compared them with others. These cases are mainly domain-related factual knowledge following the categorization of knowledge in Bloom's taxonomy [33], where knowledge is categorized into four types: factual, conceptual, procedural, and meta-cognitive, following the order of difficulties in mastering. We also noticed the existence of conceptual domain knowledge, which indicates how domain principles are applied to connect facts from data. P4, as an expert in economics, pointed out that conceptual domain knowledge plays an important role in organizing their findings. P4 mentioned that they often need to consider meta relations based on their models and theories. He often picks important data attributes to analyze and present based on the factors considered in previous theoretical models. The opinion is echoed by P6. Beyond conceptual knowledge, our participants also leverage procedural knowledge about the domain to build the meta relations following their analytical pipelines. For example, P1 mentioned that he arranged the data facts based on his commonly used strategy, where the story pieces are introduced according to the car models first and then the temporal analysis. P4 also introduced an example where

they leveraged their expertise to link the facts of the fast-increasing consumption of electricity, the relevantly slowly increasing carbon dioxide density in the atmosphere, and the increasing investment in green power to showcase the impact of such investment. We did not notice any meta-cognitive knowledge since it is about how other knowledge is learned and might be beyond our scope.

The narrative intent mainly includes the question that the story aims to answer and also the goals of the story. P3, as a designer, highlighted the impact of data story goals. She mentioned that the take-home message would affect how she organized story pieces. In our task, she noticed that two SUV models, BMW X5 and X6 had increasing sales between 2007 and 2011 while the other car models experienced a sales drop. She said that she would introduce BMW models first if the key message is that the overall SUV market is depressed. On the contrary, if the story's goal is to highlight the business success of BMW models, it is better to introduce the unsatisfactory performance of other models first to highlight BMW's unusually increasing sales.

To summarize, we have two observations about meta relations. First, unlike data relations, *meta relations have a more flexible and ambiguous nature based on users' intents and preferences*. As a result, meta relations may not have a concrete and complete categorization of their detailed meaning, such as contrast or similarity in data relations [13]. Second, though a detailed categorization of meta relations is not feasible, *our participants unanimously described the classification of meta relations using the sources where meta relations are identified*. Specifically, we identify two common sources of meta relations in data stories, *i.e.*, domain knowledge and narrative intent.

Applying meta relations is non-trivial and therefore AI assistance is expected. Though meta relations are widely applied, it is non-trivial for our participants to apply them in data storytelling. We noticed that most of the challenges are related to domain knowledge.

Several participants mentioned that their challenges are related to acquiring and digesting domain knowledge to distill meta relations (P1, P2, P3, P5, P6). When P2 conducted the tasks in our formative study, to help himself understand the content about SUV background, he fed all materials to ChatGPT, a large language model (LLM)-driven chatbot, and asked it to summarize them into bullet points. In this way, he could easily understand the differences and commonalities between SUV models. P1, P3, and P5 also found domain knowledge is often overwhelming and could be hard to digest without any assistance, which is described as "*explicit hint*" on potential meta relations by P1. This finding aligns with the common practices of data workers where they often work with diverse domain experts in different projects [7], [16], [34]. They rely on domain knowledge provided by these experts to make sense of data and interpret the findings from datasets. P2 also mentioned the difficulty of collecting domain knowledge by talking about his experience with a mobile game-related analysis project. He said "*I spent five months playing the game!*" to collect the domain knowledge about the game mechanism for the analysis task. P4 also indicated his

challenges with conceptual knowledge in the domain. He often needed to search for literature to support and verify the linkage between his results and models, where assistance was also expected. Besides domain knowledge, P3 mentioned that she might be unsure how to better compile a data story following her narrative intent. As a result, she would like to ask AI for assistance in brainstorming potential ways to tell stories or assessing the pros and cons of different approaches.

From the participants' opinions, we realized that *the main challenge lies in applying meta relations from domain knowledge* due to the unfamiliarity of a specific domain. They were keen on having approaches to help them digest domain knowledge and provide hints in potential meta relations. Furthermore, *the application of AI in addressing their challenges is observed*. P2 directly applied ChatGPT to our task, while P3 also mentioned the possibility of introducing AI to her workflow. Such observation implies the potential of introducing powerful AI models to collaborate with humans for the aforementioned issues.

IV. DESIGN CONSIDERATIONS

According to the literature and our formative study, meta relations are important in organizing story pieces into a personalized and vivid data story. This section summarizes considerations for incorporating meta relations in AI-powered data storytelling tools based on their definition.

We first formalize the definition of meta relations. In previous research (*e.g.*, [10], [11], [18]), a data relation often includes four components: two story pieces connected by the relation, the type of the relation, and a score to delineate the importance of the relation. To keep consistent with data relations, we define a meta relation $relation_{AB}$ from $fact_A$ to $fact_B$, with a relation type and an importance score of the relation, as a quadruple, $relation_{AB} := (fact_A, fact_B, type, score)$. In the definition, $fact_A$ and $fact_B$ are two data facts (*e.g.*, “Toyota Prius experience a sales performance drop” and “Nissan Leaf, Tesla S, Chevrolet Volt have growing sales” in Fig. 1). *type* is a concise description of a meta relation between two facts using the meta information of the dataset (*e.g.*, “Hybrid electric and plug-in cars are competitors in the electric car market” in Fig. 1). *score* quantifies the importance of the meta relation with a score between 0 and 1, where a larger value indicates a higher importance. Notably, there can be multiple relations between $fact_A$ and $fact_B$ with different types and corresponding scores.

The considerations include four aspects: how tools handle the definition of meta relations, what functions should be included, and what techniques could be applied. The first two considerations **C1-C2** elicit the requirements on meta relation definition.

C1. Defining meta relation types flexibly with natural language. The study results reveal an important characteristic about meta relations: *they depend on the detailed meta information and thus are flexible*. Based on the finding, it is almost impossible to categorize meta relation types into a fixed number of classes, which is common in identifying data relations with AI [13]. As a result, *the types of meta relations*

should be described with free-form natural language, inspired by a recent research [35]. In this way, it is possible to fully unlock the potential of meta relations as a medium to convey human creators' intended relations between story pieces to AI modules in these tools.

C2. Estimating the importance of meta relations in a universal way from multiple angles. When applying meta relations in composing data stories, the importance scores can be essential references for ranking multiple relations between two facts. Previously, the importance scores of data relations are often decided with specific criteria according to the relation type, such as the overlapping between attributes in adjacent facts [18]. However, it is non-trivial to use a similar approach to compute scores for meta relations due to their flexibility, as indicated in C1. Therefore, it is essential to design a universal approach to quantify the importance of meta relations with considering the sources of meta relations concurrently.

Next, **C3-C5** present *considerations about detailed functions that tools should include to address users' pain points*.

C3. Suggesting potential meta relation from users' provided domain knowledge automatically. Based on our study results and prior literature [7], [16], data workers may need to spend considerable time and effort to provide meta relations manually. Furthermore, the context switch between data analysis and information seeking is often undesired [18]. We envision that it is essential to provide an efficient way to assist data workers in distilling the meta relations between data facts from the documents provided by data workers. Inferring relations from provided documents could limit the scope of meta relation inference and ensure that the inferred ones match expectations.

C4. Following users' diverse narrative intent in the workflow. As revealed in our formative study, diverse narrative intent often guides organizing story pieces into a coherent story, such as the analytical question and the intended take-home message. Therefore, these tools should allow users to express their narrative intent freely and consider such intent when suggesting meta relations and organizing story pieces.

C5. Providing users with sufficient flexibility to select, modify, and delete automatically generated meta relations and provide customized meta relations. Compared to data relations, meta relations can be more diverse and dependent on the provided documents (C1). Therefore, the inferred meta relations may not be fully aligned with data workers' knowledge and expectations. It is crucial to provide convenient interfaces and interactions for data workers to select useful meta relations, modify their details, or reject those incorrect and redundant ones. Also, extra meta relations should be easily added.

C6. Supporting meta relations in the entire lifecycle of data storytelling. Considering the close connection between all stages when creating data stories [5], the support of meta relations should also be reflected in multiple stages.

To achieve **C1-C5**, there are several technical barriers. First, considering the free-form expression of meta relations (**C1-C2**), *we can hardly design a fixed set of algorithms to decide the relation type and compute the relation score* when recommendation (**C3**). Second, when organizing data facts

based on meta relations, *the AI should be able to interpret the semantic meaning of both these free-form meta relations and the narrative intent (C4-C5)*. Third, **C6** requires the AI system should have the ability to care for different tasks in the whole workflow. Lastly, since there is no pre-defined set of meta relation types, *it is hard to collect data and train an AI model to fulfill the needs*.

C7. Leveraging LLMs to suggest and interpret meta relations carefully. The emerging large language models (LLMs) shed light on addressing the challenges due to their advanced performance in understanding and generating natural language and the zero-shot learning ability through prompt engineering [9]. Such features allow LLMs to distill meta relations from domain knowledge and organize data facts without additional training. On the other hand, their application should be careful due to their drawbacks, *e.g.*, delay and unreliable results.

V. REMEX: EXEMPLIFYING THE DESIGN CONSIDERATIONS

To explore where and how AI should intervene when telling stories with meta relations, we designed a tool, Remex (**RElax the constraints of composing data stories with MEta relations**), that follows the design considerations. Remex integrates AI-powered computational modules to support the lifecycle of storytelling in the computational notebook, a common platform to analyze and communicate data [36] (**C6**).

A. Overview

As Fig. 2 shows, Remex is composed of five computational modules powered by AI techniques working at the backend and two interactive modules that allow humans to view and improve the output of computational modules (**C5**). Users can provide the story-related domain knowledge and their narrative intent in the text format when starting the data analysis session in computational notebooks. Then, users can add visualization specifications in Vega-Lite [29] in each cell for Remex to illustrate potentially interesting data facts using the *data fact suggestion module* and suggest meta relations with previously user-interested facts with the *meta relation identification module*. Besides, *data relation identification module* complements the meta relation identification module to infer data relations for the follow-up organization. The facts and meta relations are then presented in the *analysis panel* for users' selection and modification. When users feel satisfied with a fact with or without meta relations, they can add them to their slides for communication purposes. The fact will be inserted into a draft slide deck by the *story organization module*. The slides are shown in the *organization panel* for users' adjustment. When the analysis ends, the slide deck can be exported for follow-up improvements or direct presentation with the *slide generation module*.

B. Data Fact Suggestion

Remex leverages a data fact suggestion module to lower users' workload in documenting data findings when they

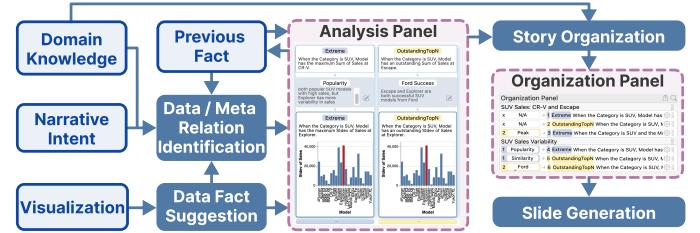


Fig. 2. This figure illustrates the structure of Remex. More details about the analysis and the organization panels are available in Fig. 5.

realize any notable findings. The module applies the well-established fact mining algorithm [12], [18], [37] to infer data facts from user-specified visualizations with seven dimensions, including a set of filters to describe how the data subset is derived (*subspace*), the independent and dependent variables in a chart (*dimension* and *measures*), and four attributes to describe the data fact (*i.e.*, the *type* of the data fact, *parameters* to indicate the fact details, the *focus* data point of the fact, and *scores* to delineate a fact's importance and alignment with users' interest). The output of this module is data facts with text descriptions and visual illustrations.

C. Data Relation Identification

The data relation identification module computes two types of data relations between data facts with the algorithm in a previous research [18].

Following the principles of sequencing story pieces [10], the first group of relations delineates the similarity between data facts through the overlap between data facts. The detailed relation types include the overlapping between subspaces, measures, dimensions, focuses, and fact types. Furthermore, the scores in such overlapping relations reflect how large the overlapping between two facts' certain attributes is. We compute it as the Intersection-Over-Union (IoU) score [38] between two facts' specific attributes, such as subspaces or measures.

The second group of relations delineates the temporal relationships between two facts and the relationships between two facts' importance scores. The temporal relationships include the temporal order of two facts' subspaces and focus. The rationale behind considering them is to introduce two facts in chronological order. The introduction of the relationships between the two facts' strengths aims to arrange the important facts before the less important ones. For these relations, we consider the relation strength score as a binary value, *i.e.*, 0 or 1. When the two facts in the relation follow the chronological order or the more important fact is before the less important one, the relation score is 1. Otherwise, the score is 0, or, in other words, the relation does not exist.

D. Meta Relation Identification

According to our formative study, AI assistance is desired by some data story creators to distill meta relations between data facts from domain knowledge (**C3**). At the same time, the AI should consider how the meta relation may fulfill the user's

narrative intent (**C4**). We design a meta relation identification engine to help users discover meta relations.

Following **C7**, we leverage LLMs to identify flexible meta relations and infer relation importance scores. Informed by their zero-shot learning ability [39], we prompt LLMs to finish three tasks, including completing the quadruples of meta relations and supporting automatic and manual verification, as shown in Fig. 3. We carefully design the two verification-related tasks to minimize unreliable results, such as hallucinated response [40], in the completion task. We feed both previously selected and new data facts, narrative intent, and domain knowledge into LLMs to conduct these three tasks.

The first task asks LLMs to *complete the quadruples of meta relations*, *i.e.*, two facts, their relation type, and the relation score. The model directly identifies the two facts and their meta relation type as a natural language description from domain knowledge (**C1**), as well as a summary for human creators to quickly understand the meta relation in one glance. Furthermore, to compute the relation score as comprehensively as possible (**C2**), we ask LLMs to self-rate the meta relation from five perspectives: *strength*, *fidelity*, *helpfulness*, *interestingness*, and *confidence*. We consider the scores estimated by LLMs to be a valid proxy for humans' perception of meta relations from two perspectives. First, most of the LLMs are trained to align with humans using various approaches, such as reinforcement learning from human feedback (RLHF) [41]. Furthermore, previous experiments have verified that LLMs often make similar judgments to humans [42] and can simulate humans in HCI experiments [43]. *Strength* provides a direct estimation of how strong the relation is. Inspired by the metrics of evaluating fact strength [12], *fidelity* evaluates how much the relation covers the domain knowledge, and *interestingness* implies whether the relation might attract story audiences. Furthermore, we also ask LLMs to consider how much the relation can help convey the narrative intent via providing a score of *helpfulness*. Finally, to reflect how reliable the result is, a *confidence* score is also returned by LLMs. All scores are returned as an integer between 1 (the lowest) to 5 (the highest), similar to the 5-point Likert scale. The final relation score is then computed based on these five scores as $score = confidence * \sum(w_i * score_i)$, where $i \in \{strength, fidelity, helpfulness, interestingness\}$. The purpose of using a weighted average of these scores is to allow users to adjust the score definition based on their requirements. In our experiments, we set all w_i s to 1 to treat all scores equally important.

Beyond the first task to return meta relations directly, the other two tasks are designed to facilitate the meta relation verification. The second task is to *facilitate the automatic verification of the identified meta relations*. The LLMs are instructed to provide meta relations' detailed entities, which can be part of the data facts, such as attributes, filters, or focuses. The task design is based on our observation that meta relations may not cover all perspectives of the fact. For example, the meta relation in Fig. 1 is mainly about the focuses of the data subspaces, *i.e.*, "Model = Toyota Prius" and "Model = Nissan Leaf, etc" belong to different types of electric cars. We ask LLMs to explicitly provide the

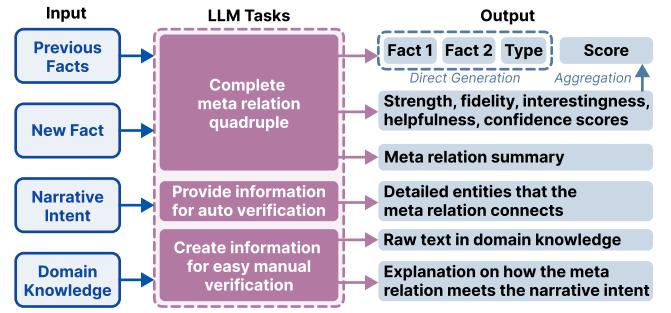


Fig. 3. The figure explains the meta relation identification module.

entity information so that we can programmatically examine whether the meta relation is a valid relation about two facts by checking the existence of entities in facts. When testing the module, we noticed that such information successfully filtered out problematic meta relations, such as relations between a domain knowledge point and a data fact.

The last task is to *facilitate the user's verification with eyeballing*. Though we have designed the second task for automatic verification, it can only handle simple issues, such as the confusion between domain knowledge and data facts. More complex issues like semantic errors in meta relations still rely on tool users to verify and resolve. Therefore, we instruct LLMs to provide evidence about their suggested meta relations from two perspectives. First, the LLM provides the original text in user-provide domain knowledge where the meta relation is derived. In this way, whenever the user feels unsure about the meta relation, they can simply refer to the original text for verification. Second, we also ask LLMs to explain how the relation is related to the narrative intent. It can serve as a hint for users to decide whether the relation should be considered in the story.

The three tasks are finally described with natural language in our prompt, together with other necessary information, including input data, meta relation definition, and rules for output (*e.g.*, output format). Due to the limited space, we include the complete prompt in our supplementary material. The whole prompt is then fed into LLMs for their responses. After receiving the responses from LLMs, we conduct a series of output processing, including output format checking, automatic meta relation verification by examining the existence of entities in facts, and meta relation ranking with importance scores.

E. Story Organization

Informed by the advantages of closely connected data analysis and storytelling [18], we would like to design Remex as a notebook extension to organize data stories along with the progress of data analysis. Therefore, the overall goal of this module is to find a suitable position for the newly added fact following its meta and data relations with previous facts. The narrative intent is also considered in the organization (**C4**). In this module, we apply LLMs in the story organization to cater to the needs of understanding the semantic meaning of meta relations and narrative intent (**C7**). Similar to the meta relation

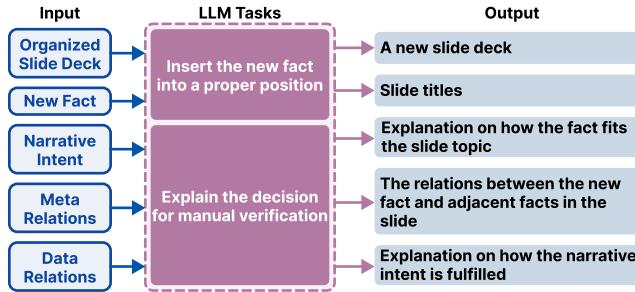


Fig. 4. The figure explains the data story organization module.

identification module, we have a main task for inserting facts into slide decks and an auxiliary task for quick verification of results. Both tasks leverage the same set of inputs, including the organized slide deck, the new fact, its data and meta relations with existing facts, and the narrative intent.

The first task of *inserting the fact into the slide deck* requires LLMs to consider the new fact’s suitable position in the slide deck from the semantic and design levels. From the semantic perspective, the model is instructed to consider how the fact fits the topic of the slide and the relations between the new fact and the facts behind or in front of it. The narrative intent of the entire story should also be carefully considered. At the slide design level, we provide two considerations, including the maximum number of facts in each slide to avoid overwhelming the story audience and the minimum alternation of the previously arranged sequence to diminish tool users’ confusion. Furthermore, a slide title is generated for each slide to summarize the content. The second task requires LLM to *provide the rationale behind the decision to insert the fact*. The rationale should also be provided to explain the four design considerations at the semantic level. In this task, we do not ask the LLM model to provide design alternatives, as multiple alternatives will increase users’ workload in reading multiple stories and selecting their preferred one. Considering that the goal of the organization module is to create a story draft during data analysis, we prefer not to increase users’ mental load in organizing story pieces. Another advantage is to reduce the LLM delay and cost.

The three tasks are then encapsulated into a prompt with the input data and requirements of output. The complete prompt is available in our supplemental material. After receiving LLM responses to the prompt, the module further processes the output. Besides sanity checking, Remex ensures LLMs do not modify user-revised content, including the fact sequences and the slide title, in this step (**C5**).

F. Interactive Modules

To facilitate users’ inspection and modification of outputs from computational modules (**C5**), Remex provides two interactive modules, including analysis panels presenting data facts and their meta relations to previously selected facts and an organization panel to show the sequence of all selected facts.

Analysis panel. The analysis panel, depicted on the left side of Fig. 5, displays the original chart on the left and

accompanying data facts on the right. Users can select from recommended options or customize their facts. Each data fact is presented in three cards. The top card (Fig. 5(a)) showcases the previous fact along with its type and description. Hovering over this card reveals the associated chart (Fig. 5(e1)). The middle card (Fig. 5(b)) illustrates the meta relation between the previous and current facts, providing a summary word and detailed description. Users can edit this description via the edit button (Fig. 5(e2)), with additional information for verification (see Sec. V-D) available upon hovering (Fig. 5(e3)). The bottom card (Fig. 5(c)) displays the current fact, including its type, description, and accompanying chart. For customized facts, users can choose from recommended options. Later users can add/remove facts as needed (Fig. 5(e4)-(e5)).

Organization panel. The organization panel, depicted on the right side of Fig. 5, enables users to outline and customize their stories. Selected data facts are displayed here, along with explanations for their placement in the dialog box (Fig. 5(e4)). In the panel, we represent each slide as a block (Fig. 5(d1)) and individual facts as list items. The encoding of a list item is introduced with a legend (Fig. 5(d2)). The left end of a list item indicates that a previously selected fact is linked with the new fact with a meta relation. It shows a summary of the meta relation and the index of the slide where the previous fact is. The color encodes the previous fact’s type for easy reference. The current fact’s details are on the right, including the index, fact type, and a text description of the fact. To export slides, users can simply click the icon as shown in Fig. 5(e6). The slide in Fig. 5(d1) will be exported to a slide as Fig. 5(d3) shows. In Fig. 5(e6), users can also hover on the magnifier icon to check the information for verifying story organization (see Sec. V-E). Users can delete or reorder facts using buttons (Fig. 5(e7)).

G. Slide Generation

This module generates a slide deck based on the organized fact sequence. Following Notable [18], there are two styles of slides to accommodate facts from the same chart or different charts. Furthermore, consider that meta relations can be a strong bond between two facts, we add an additional text box to include meta relations about two consecutive facts (see Fig. 5(d2)). We also provide the flexibility for users to customize slide background using a pre-defined set of backgrounds or through a link to the background image.

H. Implementation

In Remex, the meta relation identification and the story organization modules are supported by an LLM. After comparing the performance of GPT-4-0613 [44] and GPT-3.5-Turbo [45] provided by Microsoft Azure, we adopted the standard version of GPT-4-0613 with *temperature* = 0 inside the two modules. Notably, our approach is not limited to a specific LLM. We used the GPT family since they perform well [46] and are easy to access through Microsoft Azure APIs.

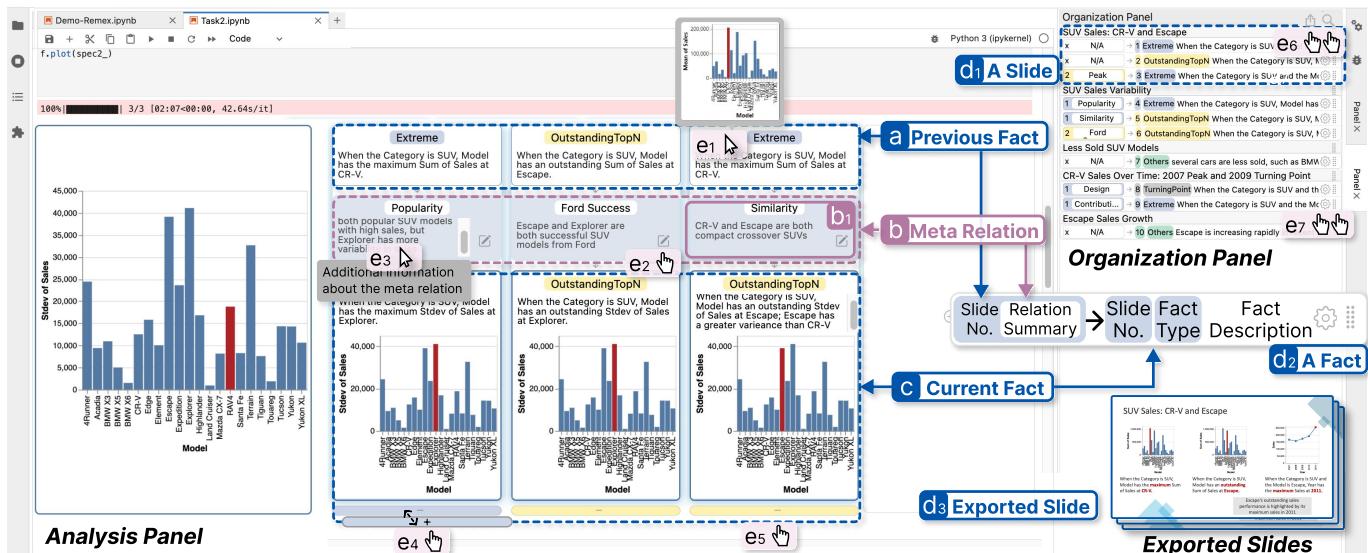


Fig. 5. This figure shows Remex in JupyterLab with a case in the user study. Remex consists of multiple analysis panels under code cells and an organization panel to show how the data findings are organized into a sequence of slides. (a)-(c) show previous facts, meta relations, and current facts in both analysis and organization panels. The arrows with (a)-(c) indicate that two linked areas show the same information. (d1) and (d2) explain a slide and a fact in the slide in the organization panel. (d3) shows exported slides. (e1)-(e7) explain the interactions. The arrow cursor indicates a hover interaction, while the hand cursor indicates a click interaction.

VI. EXPLORATORY STUDY

To explore where and how AI should provide assistance to humans with meta relations, we conducted a user study to compare Remex and a recent notebook tool without considering meta relations [18].

A. Preparation

We designed a with-in subject user study to let participants have a hands-on experience with these two tools.

Task and Dataset. In our user study, each participant was instructed to finish two tasks with different tools in a counterbalanced manner. To control the variables in the user study and minimize the memory effect, we design two tasks using two subsets from the same car sales dataset in previous research [12]. Specifically, the participants were asked to create two stories to compare different car model sales in the *Sedan* and *SUV* categories with around eight data facts. The dataset describes car sales between 2007 and 2011 with five attributes: *model*, *brand*, *category*, *sales*, and *year*. Considering the time cost of collecting domain knowledge, we provided participants with introductions to these car models from Wikipedia. The introductions often briefly describe the cars' features and histories. We also explained glossaries in car models, such as *SUV* and *MPV*, to help users understand the domain knowledge. Besides, we prepared the *cars* dataset [47] for participants to get familiar with two tools.

Participants. We invited 10 participants (5 males, 3 females, 2 prefer not to disclose, $Age_{mean} = 25.5$, $Age_{std} = 1.58$, denoted as U1-U10) from various disciplines, including computer science (U1, U4, U5, U8, U9), data science (U2, U6), social computing (U7), and business (U3, U10), with data analysis and communication backgrounds ($Experience_{mean} = 5.30$ years, $Experience_{std} = 2.06$ years) through our professional network. Similar to our formative study, we invited participants with various expertise for a more comprehensive view of our research question. Each participant received \$12.5 as compensation after the user study.

2.06 years) through our professional network. Similar to our formative study, we invited participants with various expertise for a more comprehensive view of our research question. Each participant received \$12.5 as compensation after the user study.

B. Procedure

The whole in-lab user study lasts about 100 minutes in total. After introducing the study procedure, we first collected participants' consent to record their videos and use their comments and demographic information for research purposes. Next, we conducted a tutorial session for the first tool to help participants learn how to use it, taking around 10 minutes. After the user felt confident about the tool usage, they started to use the first tool to compile a data story with assistance from the provided domain knowledge in around 30 minutes. Then, we gave the tutorial of the second tool and asked them to finish the same task with another subset of data using 40 minutes in total. Finally, we interviewed the study participants about (1) how they perceive the differences between two tools brought by the introduction of meta relations; (2) how they leverage meta relations in the user study; (3) whether the approach of integrating AI for meta relations meet their common practice and expectations; and (4) how the tool can be further improved. After finishing all interviews, the results were summarized following a similar procedure as the formative study.

C. Results

From the interview, we collected various opinions about meta relations.

Integrating meta relations into the storytelling workflow demonstrates advantages from multiple perspectives. When asking the participants about the comparison between Remex and Notable, they generally considered that meta relations

could help them with data analysis and the follow-up communication in notebooks.

We noticed that a widely recognized advantage brought by suggested meta relations is to provide context information for an in-depth understanding of data facts (U4, U5, U6, U7, U9). For example, U5 commented that Remex presents implicit information in a visible way and therefore is more “*insightful*”. Meta relation can also help participants *decide the directions of data analysis and communication* (U3, U5, U7, U8, U9). U7 mentioned that such meta relations could suggest data facts that are potentially related to the focus of the current data facts from “*high-level*” perspectives. U7 pointed out a case to support his opinion. As shown in Fig. 5(b1), where Remex presents that “CR-V and Escape are both compact crossover SUVs”. Even though U7 was not familiar with these car models, it is possible to quickly decide that the sales trends of these two models may be worth more exploration. U7 believed that there might be some interesting sales trends behind these two compact crossover SUV models since Escape has a large variance while CR-V has a larger overall sales. Similarly, the meta relation “Escape and Explorer are both successful SUV models from Ford” on the left also stimulated U7 to explore more about their sales trends for more details. Furthermore, U3 and U8 believed that Remex serves as an “*external assistant*” (U3) to help them expand the angles of data exploration with meta relations. U3 commented that “*it provides a different viewpoint to help me break the limitation of my mindset*”. Though agreeing with the benefit, U9 held a reservation about whether the benefit could surpass the potential risk of “*converging too fast in data exploration*”. U9 worried that some important data facts might not be explored due to the bias led by the suggested meta relations.

Another advantage of meta relations, as expected, is that it provides a better connection between story pieces (U5, U6, U8, U9, U10). U5 believed that the ways of organizing story pieces are enriched by introducing meta relations. U9 commented that considering meta relations could make data stories more organized by connecting data facts more closely. If only considering data relations, the story can be “*fragmented*”. The opinion was echoed by U8 and U10.

Besides, U8 also noticed that these relations, together with previous facts, could remind them of previously explored but forgotten findings. U10 considered that the explicit suggestion of meta relations could reduce the efforts of considering them mentally, especially when “*the relations between many facts form a network*”.

The expected usage scenarios of meta relations are diverse. In Remex, we provide an approach to suggest meta relations with LLMs between user-interested data facts and newly identified data facts along with their data analysis. Users can select suggested meta relations or manually add ones to share their opinions about how data facts are connected with AI for follow-up story organization. Though our current approach is appreciated for advantages like reminding users of previous facts and suggesting potential directions of analysis and storytelling, other approaches were proposed by our participants as well.

U1 mentioned that they prefer to focus on data facts instead of relations when initially exploring data. Therefore, it is not desired to directly see the recommended meta relations, which may require considerable effort to understand and increase their workload. A more desired way is to *provide a widget where two facts can be selected by users and then meta relations can be inferred and presented to users*. U1 believed this approach could help users understand the in-depth connection between two facts beyond simple data relations while minimizing users’ efforts by only showing what they expect. The idea was echoed by U3 since U3’s common workflow is to “*browse all facts first like reading menus*” and then build connections between them to tell stories. U7 also shared a similar opinion, stating that a more fine-grained interaction was expected. U7 imagined that users could select a specific data point for querying meta relations with other facts.

U5 believed that meta relations should not be limited to connecting observed facts, including both previously selected facts and the facts from the current chart. They could play a more important role in data storytelling workflow by *suggesting other potentially interesting facts associated with observed ones*. U5 took car sizes (e.g., compact or mid-size) as an example. They hoped that it was possible to recommend facts in other sizes when a meta relation related to a specific car size. In this way, users can gain a more comprehensive view of related facts. Similarly, U2 would like Remex to use meta relations to recommend what users might analyze in the next step.

U7 further proposed that *some fixed meta relations based on factual knowledge could be presented before data analysis* in notebooks (e.g., the meta relations between items in the datasets). The same idea was expressed by U2 and U5 as well. In this way, users may get opportunities to become familiar with the meta relations prior to data analysis rather than learning new meta relations continuously during data analysis. Furthermore, as U2 pointed out, such a pre-processing step might help reduce the time cost of LLMs in the follow-up meta relation recommendation.

The display of meta relations is expected to be simple and effective. Since meta relations connect two data facts, when displaying meta relations to users, they require users not only to understand the relations themselves but also to recall an existing fact and to understand a new data fact. Such an issue adds considerable cognitive load to users during the analysis. As a result, the display of meta relations needs to be carefully designed to reach a balance between information size and the extra workload for users to understand it. In our study, several users feel satisfied with the current design of meta relation display in Remex (U6, U8, U9, U10). For example, U10 commented “*the combination of charts and texts is nice*”. Besides, we collected multiple suggestions on the display of meta relations to facilitate understanding.

U1 preferred to show meta relations only when users query. U4 considered the current design to be “*unnecessarily cumbersome*” since users need to read every meta relation. To address the issue, U4 believed that a potential approach was to provide better summarizations of data facts and meta relations to users. In U4’s opinion, the current one-word summary

was not informative, while the complete meta relation was lengthy. An ideal summarization of U4 would be using one short sentence to describe each fact and another one to show the meta relation between them. Another alternative approach would be showing a synthesized chart that introduces the two facts and the meta relation between them at once. U7 described another relation-centric approach. When users plot new charts, only meta relations might be shown initially so that users can have a quick glance. When meta relations interest users, they may expand the facts for a detailed inspection and decide if they will be included in the story.

Compared to domain knowledge, narrative intent is more dynamic. We observed that users heavily relied on meta relations from domain knowledge, while the role of narrative intent was less notable. At the beginning of the analysis, most of the users did not have a clear intent about the story they wanted to tell. Their narrative intent is gradually built and changed after they learn more about data. For example, U6 noticed that the top-selling sedan for multiple years remained consistent, but its competitors changed a lot. Then he decided to introduce the sales of cars year by year. U7 noticed that some car models obviously have a greater standard deviation in sales and then decided to focus on the trend of car sales to explain the significant sales variance over the years. Though Remex allows users to update their narrative intent whenever they want, we noticed that our study participants did not use the function. We consider the main reason might be that users find it cumbersome to update it every time the intent is changed. For example, U2 stated that they preferred an AI to follow their intent implicitly during the usage.

VII. DISCUSSION

This section discusses the lessons we learned, the limitations of our work, and future work.

A. Lessons

We learned three important lessons throughout our research project about applying meta relations in AI-powered data storytelling.

Making invisible meta relation sensible by creators and AI. Our research was motivated by the wide application of meta relations in practices, such as the example from Bloomberg (Fig. 1). However, meta relations were not reflected in the design of AI-powered storytelling tools, which led to the gap between humans' expected data stories and those created by AI. One potential reason is that diverse meta relations are often implicitly considered by data story creators, like other meta information in visualization [15]. As a result, the importance of meta relations for composing data stories is often overlooked. In this paper, we conducted research to explore meta relations in data stories from several angles, including their categorization and characteristics in practice and the timing and approach of AI assistance.

During our research, we also realized an unresolved challenge in making humans and AI aware of meta relations. Between the two sources of meta relations, *i.e.*, domain knowledge and narrative intent, story creators and AI are aware

of meta relations from domain knowledge easily. We consider the reasons to be two-fold. First, *domain knowledge is almost static and objective while narrative intent is dynamic and personalized*. Therefore, the synchronization between humans and AI in domain knowledge is convenient without much cost. It is sufficient to provide domain knowledge once before the collaborative storytelling between humans and AI. However, communicating narrative intent would be more challenging. Story creators need to continuously update their intent with AI. Furthermore, describing the vague intent with natural language when communicating with AI can be hard. We found that Remex's function of updating narrative intent has almost never been used in our user study. Second, *the meta relations led by domain knowledge are mostly directly reflected by the connection between two facts, but narrative intent may not have a strong impact on a specific meta relation*. Instead, narrative intent may affect the importance of meta relations. For example, when the narrative intent is "telling a story about luxury cars", a meta relation "Mercedes-Benz S-Series and BMW 7-Series are often selected by millionaires" should be more important than "Mercedes-Benz S-Series and BMW 7-Series are both cars from German carmakers." If the intent is "telling a story about European cars", then the relation about German carmakers is more likely to fit the story. As the case shows, narrative intent may not be directly reflected in the relation descriptions, making it less sensible. We consider that future research should delve into the challenge led by narrative intent, including how to facilitate effective communication of narrative intent between human creators and AI and how to make the impact of narrative intent more visible through visual design.

Designing tools for applying meta relations in personalized ways. Our formative and user study informed the importance of offering sufficient personalization when designing AI-powered storytelling tools for meta relations. We realized three levels of personalization in tool designs: *meta relation source, usage scenario, and display*.

According to our observations, *source* preference varies among story creators. Some preferred to have strong and clear narrative intent before creating stories (*e.g.*, P3), while others constructed their narrative intent gradually by identifying relations between story pieces (*e.g.*, U6, U7). Considering categories of domain knowledge, we noticed that some users might use factual knowledge more frequently, while others considered conceptual and procedural knowledge more important.

After suggesting meta-relations based on users' mindsets, the suggestions should be delivered using four approaches in different *usage scenarios*. The first approach is to recommend meta relations when two facts are fixed. It requires the lowest load of understanding meta relations but users also lose opportunities to identify potentially interesting meta relations beyond the selected facts. The second one is to fix one fact and recommend multiple potential facts and meta relations between them. The approach can help users decide the next step of analysis or storytelling with the assistance of meta relations. The drawback is that it requires considerable computation to decide the following facts due to the large

number of facts in datasets. One potential solution is to leverage the Monte Carlo Tree Search [13], [14], [21]. The third method is to find meta relations between previously selected facts and a group of user-focused facts, as Remex does. This approach balances the aforementioned approaches from the computational cost perspective since the space of facts is bounded but not limited to only two. Users can also get unexpected meta relations to inspire their story, as our participants suggest (see Sec. VI-C). However, it indeed requires users to understand more meta relations than the first approach, which is less favorable by some participants. The last approach is to show meta relations before analysis to provide background information to users, similar to a mind map of factual knowledge to guide storytelling. However, since new data facts are spotted in data analysis, the approach may not cover rich meta relations between newly discovered facts. We would like to encourage future research to explore other potential usage scenarios (*e.g.*, identifying facts with specific meta relations) and investigate how to combine multiple types of assistance based on personal preferences.

Finally, even if we can suggest suitable meta relations in a desired scenario, the *display* of meta relations is also worth careful consideration. According to our study, some users suffered from information overload while others did not. Such a difference requires us to design interactive displays of meta relations to cater to different needs.

Applying an effective combination of LLMs and rule-based approaches for considering meta relations. Compared to data storytelling tools based on facts and relations from data [12], [13], the tools with meta relations require interpreting and suggesting natural language-based meta relations with LLMs. However, LLMs are not without limitations. For example, their high delay has affected the user experience with Remex (U5, U8, U9, U10). To help readers understand the issue, we report the time spent identifying meta relations (Min: 25, Max: 80, Mean: 47, Std: 16) and organizing data facts (Min: 22, Max: 56, Mean: 37, Std: 13) in seconds with GPT-4 in U5's session. Sometimes the identified meta relations are unreasonable or not in-depth. For example, U2 noticed that some meta relations returned by LLMs were only a repetition of two facts (*e.g.*, a suggested meta relation between a fact about increasing sales of BMW and another indicating Toyota as the brand with highest sales is “BMW Sales decrease while Toyota has the highest sales”). U8 commented that the background information provided by meta relations was not out of expectation. Furthermore, LLM's ability to understand and analyze data is limited [48]–[50]. Their bias [51] and hallucinations [52] even damage the credibility.

To address these issues, an effective combination of heuristics-based approaches with LLMs is adopted. We consider heuristics-based approaches to be more deterministic but may not be as capable as LLMs. Therefore, they can be leveraged to enhance the reliability and accuracy of AI-powered tools, while LLMs can provide more flexibility. For example, in Remex, we leverage rule-based approaches to compute data facts and apply LLMs to handle meta relations. We also apply rule-based automatic verification for LLM-generated content, such as checking the integrity of the two

facts connected by a meta relation. Moreover, the combination of algorithms can be enhanced in future tools. For example, the delay led by LLMs might be reduced by distilling meta relations from factual knowledge prior to the entire storytelling workflow or introducing external knowledge bases [53].

Beyond combining LLMs and rule-based approaches, a combination of LLMs and other deep learning methods is worth exploring. For example, inspired by prior research [42], [54], we might employ another LLM to check the suggested meta relations and organized data story, considering the inability of rule-based methods. Besides, U2 pointed out that a large quantity of domain knowledge could enhance the usefulness of Remex but worried about the scalability. Retrieval-augmented generation (RAG) [55] might be applied to address the issue.

B. Limitations and Future Work

As a preliminary investigation, we admit that our work has limitations and hope the understanding of meta relations can be enhanced in future.

The limitation of our research mainly lies in the design of our user study. First, considering the time limit, we provided users with a relatively small dataset and asked them to create a short story following prior research [12], [18]. U7 believed that meta relations could have demonstrated their values more clearly when the dataset was more complex. Future research should conduct more field studies or longitudinal user studies to observe the practices of meta relations for more insights. Second, the scale of our exploratory study was limited to ten data workers. We believe it will be beneficial if the scale could be extended with the involvement of other potential users, such as data journalists and researchers in natural science. A larger experiment can reveal more insights regarding where and how AI could support meta relation usages, such as more scenarios and interface design preferences. We hope future research could investigate the topic with more participants to substantiate our findings. Third, Remex only covers limited cases of data storytelling. Remex serves for telling stories with slideshows, one of the most common formats of data stories [10]. However, data storytelling covers a broad spectrum beyond slideshows [56], such as telling stories as data articles [57], [58] or even in virtual reality or augmented reality environments [59], [60]. It is interesting to explore whether our findings can be generalized.

We are also interested in various future directions to enhance our research. First, we hope to further enhance the theoretical foundation of meta relations. For example, it will be interesting to investigate approaches to categorize meta relations other than their sources and propose design spaces for future usage. Second, Remex can be enhanced from multiple perspectives. For example, more formats of domain knowledge, *e.g.*, figures, can be supported by Remex. It will also be interesting to enhance Remex's compatibility with other software, such as video editing tools, to support data storytellers' practices.

VIII. CONCLUSION

Meta relations provide a unique perspective for data storytellers to connect various story pieces into a coherent and

vivid story. However, they have not been thoroughly examined and considered in AI-powered data storytelling, leaving a gap between AI-created data stories and humans' expectations. To address the issue, we conducted a formative study to understand story creators' practices regarding meta relations, such as common meta relation sources and their pain points about meta relations. Next, based on the design consideration learned from the formative study, we deepen the understanding of where and how AI should take actions in the storytelling workflow for meta relations through a user study with a proof-of-concept tool, Remex. The user study revealed a series of findings about applying meta relations in the data storytelling workflow, including the diverse benefits and various expected usage scenarios. As an initial step towards introducing meta relations into AI-powered storytelling tools, we also learned multiple precious lessons. We hope future research could enhance the theoretical foundation of meta relations and their applications in practice.

REFERENCES

- [1] H. Li, "Bridging data analysis and storytelling with human-ai collaborative tools," Ph.D. dissertation, The Hong Kong University of Science and Technology, 2024.
- [2] F. Yang, M. Cai, C. Mortenson, H. Fakhari, A. D. Lokmanoglu, J. Hullman, S. Franconeri, N. Diakopoulos, E. C. Nisbet, and M. Kay, "Swaying the public? Impacts of election forecast visualizations on emotion, trust, and intention in the 2022 us midterms," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 01, pp. 23–33, 2024.
- [3] M. Brehmer and R. Kosara, "From jam session to recital: Synchronous communication and collaboration around data in organizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 1139–1149, 2021.
- [4] N. W. Kim, H. Im, N. Henry Riche, A. Wang, K. Gajos, and H. Pfister, "Dataselfie: Empowering people to design personalized visuals to represent their data," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 79:1–79:12.
- [5] 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*, vol. 35, no. 5, pp. 84–90, 2015.
- [6] K. Schröder, W. Eberhardt, P. Belavadi, B. Ajadilish, N. van Haften, E. Overes, T. Brouns, and A. C. Valdez, "Telling stories with data—a systematic review," *arXiv preprint arXiv:2312.01164*, 2023.
- [7] 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.
- [8] Q. Chen, S. Cao, J. Wang, and N. Cao, "How does automation shape the process of narrative visualization: A survey of tools," *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [9] H. Li, Y. Wang, and H. Qu, "Where are we so far? Understanding data storytelling tools from the perspective of human-ai collaboration," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024.
- [10] J. Hullman, S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar, "A deeper understanding of sequence in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2406–2415, 2013.
- [11] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer, "Graphscape: A model for automated reasoning about visualization similarity and sequencing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2628–2638.
- [12] 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*, vol. 26, no. 1, pp. 895–905, 2020.
- [13] 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*, vol. 27, no. 2, pp. 453–463, 2020.
- [14] G. Wu, S. Guo, J. Hoffswell, G. Y.-Y. Chan, R. A. Rossi, and E. Koh, "Socrates: Data story generation via adaptive machine-guided elicitation of user feedback," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 01, pp. 131–141, 2024.
- [15] A. Burns, C. Lee, T. On, C. Xiong, E. Peck, and N. Mahyar, "From invisible to visible: Impacts of metadata in communicative data visualization," *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [16] H. Lin, M. Lismic, D. Akbaba, M. Meyer, and A. Lex, "Here's what you need to know about my data: Exploring expert knowledge's role in data analysis," 2023.
- [17] A. Pearce, B. Migliozzi, and D. Ingold, "Scientific proof that americans are truck addicts," <https://www.bloomberg.com/graphics/2015-autosales/>, Jan 2015.
- [18] 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*, 2023, pp. 173:1–173:16.
- [19] F. Chevalier, M. Tory, B. Lee, J. van Wijk, G. Santucci, M. Dörk, and J. Hullman, "From analysis to communication: Supporting the lifecycle of a story," in *Data-Driven Storytelling*. AK Peters/CRC Press, 2018, pp. 151–183.
- [20] J. Lu, W. Chen, H. Ye, J. Wang, H. Mei, Y. Gu, Y. Wu, X. L. Zhang, and K.-L. Ma, "Automatic generation of unit visualization-based scrolltelling for impromptu data facts delivery," in *Proceedings of the 2021 IEEE Pacific Visualization Symposium*. IEEE, 2021, pp. 21–30.
- [21] 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*, vol. 29, no. 1, pp. 983–993, 2022.
- [22] J. E. Zhang, N. Sultanum, A. Bezerianos, and F. Chevalier, "Dataquilt: Extracting visual elements from images to craft pictorial visualizations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.
- [23] W. Cui, X. Zhang, Y. Wang, H. Huang, B. Chen, L. Fang, H. Zhang, J.-G. Lou, and D. Zhang, "Text-to-viz: Automatic generation of infographics from proportion-related natural language statements," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 906–916, 2019.
- [24] L. Ying, Y. Wang, H. Li, S. Dou, H. Zhang, X. Jiang, H. Qu, and Y. Wu, "Reviving static charts into live charts," *arXiv preprint arXiv:2309.02967*, 2023.
- [25] L. Shen, Y. Zhang, H. Zhang, and Y. Wang, "Data player: Automatic generation of data videos with narration-animation interplay," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, p. 109–119, 2023.
- [26] B. Lee, R. H. Kazi, and G. Smith, "Sketchstory: Telling more engaging stories with data through freeform sketching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2416–2425, 2013.
- [27] B. D. Hall, L. Bartram, and M. Brehmer, "Augmented chironomia for presenting data to remote audiences," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022, pp. 1–14.
- [28] L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao, "A design space for applying the freytag's pyramid structure to data stories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 922–932, 2022.
- [29] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vegalite: A grammar of interactive graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 341–350, 2016.
- [30] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2231–2240, 2011.
- [31] A. X. Zhang, M. Muller, and D. Wang, "How do data science workers collaborate? Roles, workflows, and tools," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW1, pp. 1–23, 2020.
- [32] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson, "How data science workers work with data: Discovery, capture, curation, design, creation," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, 2019, pp. 1–15.
- [33] M. Forehand, "Bloom's taxonomy: Original and revised," *Emerging Perspectives on Learning, Teaching, and Technology*, vol. 8, pp. 41–44, 2005.
- [34] D. Piorkowski, S. Park, A. Y. Wang, D. Wang, M. Muller, and F. Portnoy, "How ai developers overcome communication challenges in a multidisciplinary team: A case study," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–25, 2021.

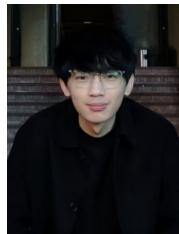
- [35] Z. Yan, C. Yang, Q. Liang, and X. Chen, “Xcreation: A graph-based crossmodal generative creativity support tool,” in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 48:1–48:15.
- [36] A. Rule, A. Tabard, and J. D. Hollan, “Exploration and explanation in computational notebooks,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 32:1–32:12.
- [37] R. Ding, S. Han, Y. Xu, H. Zhang, and D. Zhang, “Quickinsights: Quick and automatic discovery of insights from multi-dimensional data,” in *Proceedings of the 2019 ACM International Conference on Management of Data*, 2019, pp. 317–332.
- [38] H. Rezatofighi, N. Tsai, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA: IEEE Computer Society, 2019, pp. 658–666.
- [39] J. Wei, M. Bosma, V. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, “Finetuned language models are zero-shot learners,” in *Proceedings of the International Conference on Learning Representations*, 2021.
- [40] P. Lee, S. Bubeck, and J. Petro, “Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine,” *New England Journal of Medicine*, vol. 388, no. 13, pp. 1233–1239, 2023.
- [41] Y. Wang, W. Zhong, L. Li, F. Mi, X. Zeng, W. Huang, L. Shang, X. Jiang, and Q. Liu, “Aligning large language models with human: A survey,” *arXiv preprint arXiv:2307.12966*, 2023.
- [42] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. Xing *et al.*, “Judging llm-as-a-judge with mt-bench and chatbot arena,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [43] P. Hämäläinen, M. Tavast, and A. Kunnari, “Evaluating large language models in generating synthetic hci research data: a case study,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–19.
- [44] OpenAI, “GPT-4 technical report,” *arXiv preprint arXiv:2303.08774*, 2023.
- [45] ———, “GPT-3.5 Turbo,” <https://platform.openai.com/docs/models/gpt-3-5-turbo>.
- [46] W.-L. Chiang, L. Zheng, Y. Sheng, A. N. Angelopoulos, T. Li, D. Li, H. Zhang, B. Zhu, M. Jordan, J. E. Gonzalez *et al.*, “Chatbot arena: An open platform for evaluating llms by human preference,” *arXiv preprint arXiv:2403.04132*, 2024.
- [47] Vega, “Vega/vega-datasets: Common repository for example datasets used by vega-related projects,” <https://github.com/vega/vega-datasets>.
- [48] Z. Jin, X. Jiang, X. Wang, Q. Liu, Y. Wang, X. Ren, and H. Qu, “Nunsgpt: Improving numeracy ability of generative pre-trained models,” in *Proceedings of the International Symposium on Large Language Models for Financial Services@ IJCAI*, 2023.
- [49] L. Cheng, X. Li, and L. Bing, “Is gpt-4 a good data analyst?” in *Findings of the Association for Computational Linguistics: EMNLP*. ACL, 2023, pp. 9496–951.
- [50] A. Narayan, I. Chami, L. Orr, and C. Ré, “Can foundation models wrangle your data?” *Proceedings of the VLDB Endowment*, vol. 16, no. 4, pp. 738–746, 2022.
- [51] E. Ferrara, “Should chatgpt be biased? Challenges and risks of bias in large language models,” *arXiv preprint arXiv:2304.03738*, 2023.
- [52] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, L. Wang, A. T. Luu, W. Bi, F. Shi, and S. Shi, “Siren’s song in the ai ocean: a survey on hallucination in large language models,” *arXiv preprint arXiv:2309.01219*, 2023.
- [53] Y. Wang, J. Lin, Z. Yu, W. Hu, and B. F. Karlsson, “Open-world story generation with structured knowledge enhancement: A comprehensive survey,” *Neurocomputing*, p. 126792, 2023.
- [54] S. Shankar, H. Li, P. Asawa, M. Hulsebos, Y. Lin, J. Zamfirescu-Pereira, H. Chase, W. Fu-Hinthon, A. G. Parameswaran, and E. Wu, “Spade: Synthesizing assertions for large language model pipelines,” *arXiv preprint arXiv:2401.03038*, 2024.
- [55] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. tau Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-augmented generation for knowledge-intensive nlp tasks,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [56] E. Segel and J. Heer, “Narrative visualization: Telling stories with data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.
- [57] 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*, 2021, pp. 16:1–16:17.
- [58] 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*, vol. 28, no. 1, pp. 184–194, 2021.
- [59] L. Yang, A. Wu, W. Tong, X. Xu, Z. Wei, and H. Qu, “Understanding 3d data videos: From screens to virtual reality,” in *Proceedings of the 2023 IEEE 16th Pacific Visualization Symposium*. IEEE, 2023, pp. 197–206.
- [60] W. Tong, K. Shigyo, L.-P. Yuan, M. Fan, T.-C. Pong, H. Qu, and M. Xia, “Vistellar: Embedding data visualization to short-form videos using mobile augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, 2024.



Haotian Li is currently a Ph.D. candidate in Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). His main research interests are data visualization, visual analytics, and human-computer interaction. He received his BEng in Computer Engineering from HKUST. For more details, please refer to <https://haotian-li.com/>.



Lu Ying is currently a Ph.D. candidate at the State Key Lab of CAD&CG, Zhejiang University. Her main research interests are on data storytelling, glyph-based visualization. She is dedicated to integrating the AI technique into visualization to ease the creation of visualization. She received her BEng in Digital Media Technology from Zhejiang University. For more details, please refer to <https://yiyinyingu.github.io/>.



Leixian Shen is a PhD student in the Department of Computer Science and Engineering at The Hong Kong University of Science and Technology. He received his master’s degree in Software Engineering from Tsinghua University in 2023 and obtained his bachelor’s degree in Software Engineering from Nanjing University of Posts and Telecommunications in 2020. His research interests include visual data analysis and storytelling. For more details, please refer to <https://shenleixian.github.io/>.



Yun Wang is a senior researcher in the Data, Knowledge, Intelligence (DKI) Area at Microsoft. Her research lies in the intersection of Human-Computer Interaction, and Information Visualization. She develops innovative techniques and interactive systems to facilitate Human-AI Collaboration, Human-Data Interaction, Visual Communication, and Data Storytelling through an interdisciplinary approach. She received her Ph.D. from the Hong Kong University of Science and Technology. For more details, please refer to <https://www.microsoft.com/en-us/research/people/wangyun/>.



Yingcai Wu is a Professor at the State Key Lab of CAD&CG, Zhejiang University. His main research interests are information visualization and visual analytics, with focuses on urban computing, sports science, immersive visualization, and social media analysis. He received his Ph.D. degree in Computer Science from The Hong Kong University of Science and Technology. Prior to his current position, Dr. Wu was a postdoctoral researcher in the University of California, Davis from 2010 to 2012, and a researcher in Microsoft Research Asia from 2012 to 2015. For more information, please visit <http://www.ycwu.org>.



Huamin Qu is a chair professor in the Department of Computer Science and Engineering (CSE) at the Hong Kong University of Science and Technology (HKUST) and also the director of the interdisciplinary program office (IPO) of HKUST. He obtained a BS in Mathematics from Xi'an Jiaotong University, China, an MS and a PhD in Computer Science from the Stony Brook University. His main research interests are in visualization and human-computer interaction, with focuses on urban informatics, social network analysis, E-learning, text visualization, and explainable artificial intelligence (XAI). For more information, please visit <http://huamin.org/>.