Int. J. Human–Computer Studies 205 (2025) 103655

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

StoryPoint: GenAI-supported domain-specific data story authoring for enterprises

Jonas Gunklach *, Elias Mueller, Merlin Knaeble, Alexander Maedche

Karlsruhe Institute of Technology (KIT), Human-Centered Systems Lab (h-lab), Karlsruhe, Germany

ARTICLE INFO

Keywords:
Data storytelling
Authoring
Visualization
Domain expert
Enterprises

ABSTRACT

In today's data-driven world, enterprises face the dual challenge of deriving value from vast datasets and effectively communicating insights. While data storytelling is integral for conveying insights, existing authoring tools often fail to fully leverage domain expertise or support the entire storytelling process. This paper introduces StoryPoint, an open-source data story authoring tool that combines domain expertise with Generative AI to dynamically enhance data story creation. We found that enabling domain experts to intuitively visualize data via natural language inputs, supported by automatically generated charts and narratives, helps narrow the gap between visualization and interpretation. To design StoryPoint, we followed a literature-grounded and user-centered design approach. A formative evaluation with eight domain experts reveals StoryPoint's efficacy in rapid data story prototyping. A summative evaluation, including (i) a small-scale experiment comparing StoryPoint with a benchmark, (ii) a large-scale online experiment with 104 crowdworkers, and (iii) three real-world industry cases, underscores its utility. Our findings highlight that StoryPoint reduces data story creation time by more than 50% compared to the benchmark, receives significantly higher usability ratings, and supports the creation of data stories rating higher in readability, fluency, clarity, and trustworthiness.

1. Introduction

Nowadays, the utilization of data for informed decision-making and gaining valuable insights has become a cornerstone for supporting daily business operations (Trieu et al., 2022). Consequently, the abundance of data presents both an opportunity and a challenge, underscoring the need to derive meaningful value from it (Abbasi et al., 2016). However, the visualization of data alone does not inherently deliver value for users - the insights contained must be communicated suitably to the appropriate users in the right format (Watson, 2017; Gunklach et al., 2024). Recent research has shown that individuals who rely on data to make decisions often find it difficult to interpret and understand the information presented (Lennerholt et al., 2021). This underscores the growing significance of data storytelling as a means of presenting data-driven information (Boldosova and Luoto, 2019). Data storytelling is an approach for conveying data-driven insights using narratives in combination with visualizations that engage audiences and supports the efficiency and effectiveness of information retrieval and insight comprehension (Shao et al., 2024).

In enterprises, data storytelling has emerged as a transformative practice that enhances decision-making across operational, tactical, and strategic levels (Amini et al., 2018; Ramm et al., 2021). Unlike conventional dashboards, data stories aim to provide contextualized insights to align with specific organizational decision-making goals (Gunklach et al., 2023). Tory et al. (2022) identify two distinct ways enterprises rely on data to make these decisions: conversations *with* data and conversations *around* data. Conversations with data involve straightforward tasks such as monitoring KPIs, tracking metrics, or identifying trends (typically on the operational level). In contrast, conversations around data address complex, abstract problems that require synthesis of data including contextual understanding such as deciding to enter a new geographical market or invest in the development of an innovative product (on the tactical and strategic levels). Empirical research highlights that well-constructed data stories streamline operations, enhance strategic planning, and foster a culture of data-driven decision-making by aligning employees with organizational goals and improving insight comprehension through narrative sensemaking (Kandel et al., 2012; Boldosova and Luoto, 2019, 2020). Moreover, integrating data storytelling techniques, such as explanatory elements and targeted visual cues, significantly improves the accuracy and efficiency of data interpretation, making insights more actionable and relevant in diverse

* Corresponding author.
E-mail addresses: jonas.gunklach@kit.edu (J. Gunklach), udjks@student.kit.edu (E. Mueller), merlin.knaeble@kit.edu (M. Knaeble), alexander.maedche@kit.edu (A. Maedche).

https://doi.org/10.1016/j.ijhcs.2025.103655
Received 16 February 2025; Received in revised form 24 September 2025; Accepted 3 October 2025
Available online 15 October 2025

organizational contexts (Gunklach et al., 2023; Shao et al., 2024). Despite these advantages, enterprises face persistent challenges in implementing data storytelling effectively, including aligning insights with decision-making objectives (Tory et al., 2022), managing collaboration across roles in the storytelling process (Crisan et al., 2020), and addressing the domain-specific nature of storytelling (Gunklach et al., 2023).

Crafting effective data stories is inherently complex and highly creative, requiring not only technical expertise but also a deep understanding of organizational context and stakeholder needs (Zhao et al., 2023; Hu et al., 2018). This process involves identifying patterns, structuring narratives, and selecting appropriate visualizations for the target audience (Lee et al., 2015; Chevalier et al., 2018) — tasks made more challenging by the political and contextual dynamics that influence how data is framed and interpreted (Baumer et al., 2022). The storytelling process further necessitates collaboration among multiple roles: analysts are skilled at data analysis but often lack domain-specific knowledge, while domain experts provide contextual insights but in practice regularly lack technical proficiency (Crisan et al., 2020). Hence, these roles form bottlenecks due to frequent feedback loops that slow the process and risk loss of critical insights (Lennerholt et al., 2021). Despite the availability of existing data storytelling tools such as Calliope (Shi et al., 2021) or DataShot (Wang et al., 2020), significant gaps persist for enterprise data storytelling. Recent research shows that most tools fail to integrate domain knowledge effectively, resulting in generic outputs that require manual refinement, and lack comprehensive support across all storytelling stages, focusing instead on isolated phases (Li et al., 2024). Additionally, static outputs, the absence of iterative feedback mechanisms, and limited integration with enterprise systems hinder adaptability and scalability for enterprise contexts (Ren et al., 2023).

Recognizing these limitations, recent advances in Generative AI (GenAI) have sought to address parts of this challenge by automating elements of the storytelling process (Chen et al., 2023). Tools like Tableau's Pulse and Microsoft's Copilot for Power BI leverage conversational analytics to automate querying, narrative generation, and visualization, enabling users to interact with data more intuitively (Tableau, 2025; Microsoft Corporation, 2025). However, while GenAI excels at automating repetitive tasks, it cannot replace the nuanced understanding and contextual knowledge that domain experts bring to storytelling (Hutchinson et al., 2024; He et al., 2024). Generic AI-generated narratives and visualizations often require significant adaptation to align with business-specific goals, limiting their effectiveness as standalone solutions (Jung et al., 2025; Holzinger et al., 2022). Research has shown that domain-specific AI systems outperform more general-purpose models, highlighting the value of embedding domain expertise into tool design (Logeshwaran et al., 2024). This motivates our research to explore the following research questions: *(1) What design principles should data visualization tools follow to effectively support domain experts in the data storytelling process? (2) How effective are these design principles in improving the efficiency and effectiveness of domain experts in creating data stories?*

To answer these questions, we conducted a human-centered design study relying on two complementary development approaches: (1) a rigorous literature-grounded approach, where we conducted a systematic literature review on data storytelling, and (2) a user-centered design approach, where we conducted twelve semi-structured interviews with domain experts. We combined both to carefully derive requirements and design principles, based upon which we built a prototype, namely *StoryPoint*, to test different design hypotheses with potential users. To determine the impact of the first prototypical instantiation of *StoryPoint* on domain experts' ability to create compelling data stories, we started with a formative evaluation with eight domain experts using think-aloud sessions. On this basis, we refined *StoryPoint* and finally conducted a summative evaluation, encompassing a small-scale experiment with domain experts, a large-scale online experiment with 104 crowdworkers, and real-world data visualization use cases to assess the efficiency and effectiveness of its design principles. We found that domain experts were able to create quick data story prototypes with the help of *StoryPoint* by relying on natural language input, pre-calculated charts, and automatically generated narratives.

Our work has three main contributions: First, through a comprehensive requirements elicitation process conducted in the form of a systematic literature review and 12 interviews, we formulate design principles that provide a solid foundation for supporting data storytelling in enterprises. Second, with *StoryPoint* we introduce the first open source[1] data story authoring tool. *StoryPoint* introduces a novel approach that enables the integration of domain expertise into data storytelling workflows. By adapting the LLMs used for narrative and chart creation to include contextual metadata and user-defined story purposes, *StoryPoint* addresses the persistent challenge of creating domain-specific narratives and visualizations. This ensures that domain expertise is actively embedded into the storytelling process, addressing both technical and contextual challenges. By making our tool freely available to the public, we aim to enhance accessibility, foster future research, and continually improve data storytelling practices. Third, we empirically demonstrate *StoryPoint*'s utility by following a rigorous research and engineering process that includes a formative evaluation with eight domain experts and a subsequent summative evaluation. The summative evaluation comprises three parts: (i) a small-scale experiment with domain experts involving two different tasks in which *StoryPoint* is systematically compared to a benchmark, (ii) a large-scale online experiment involving 104 crowdworkers from Prolific, in which the quality of the data stories generated with *StoryPoint* is evaluated against those of the benchmark, and (iii) real data visualization use cases with authentic enterprise data that demonstrate the versatility and capabilities of *StoryPoint* in implementing different use cases. Additionally, in our evaluation we introduced scales for assessing narrative elements and data story quality - these scales not only provide robust metrics for evaluating StoryPoint but can also serve as a valuable framework for future research on data story quality across different tools and contexts. Lastly, for practical implications, we propose that StoryPoint's approach to integrating domain expertise into GenAI-supported data storytelling can inspire new directions for enterprise tools. Our intention is to empower organizations to leverage data storytelling effectively, facilitating the creation of compelling narratives around their data. In conclusion, our work contributes not only to the advancement of data storytelling but also to the broader landscape of data visualization. By providing the open source data story authoring tool *StoryPoint* and the opportunity to instantiate our design principles in commercial data visualization tools, we strive to make meaningful strides toward improved data literacy, defined by the OECD as "the ability to derive meaningful information from data [...] including how to read charts appropriately" (OECD, 2020), thus contributing to the OECD Learning Compass 2030, empowering learners to navigate and create in an increasingly data-rich world (OECD, 2019).

2. Background and related work

In this section, we first provide the necessary background on the role of data storytelling in enterprises before reviewing prior tools that informed the design of *StoryPoint*. We further discuss the recent emergence of GenAI for storytelling and human–AI collaboration, and conclude by linking these areas of work to our research questions.

[1] Source code and live demo of StoryPoint: https://github.com/Eliasm001/StoryPoint/tree/main or story-point.de.

2.1. Background: Data storytelling in the enterprise context

Data storytelling in enterprises is a transformative approach that integrates data visualization, narrative construction, and domain expertise to enhance decision-making and foster understanding across organizational levels (Showkat and Baumer, 2021). Following the seminal work of Segel and Heer (2010), this interdisciplinary practice has been recognized as a powerful technique for addressing the challenges inherent in complex, data-rich environments. Research highlights its benefits: Kandel et al. (2012) show that effective data stories streamline operations, uncover inefficiencies, and improve strategic planning. Boldosova and Luoto (2019) emphasize its role in fostering a data-driven culture and aligning employees with organizational goals. Furthermore, Boldosova and Luoto (2020) argue that storytelling improves data interpretation and decision-making by engaging stakeholders emotionally and cognitively, increasing relevance and utility. Gunklach et al. (2023) highlight how narrative elements improve transparent interaction with the data story and ensure insights are actionable, while Shao et al. (2024) demonstrate that embedding storytelling elements into visualizations enhances data comprehension and retrieval accuracy. However, despite these benefits, enterprises face challenges in implementing storytelling effectively.

One of the most significant is the need for **collaboration across roles along the data storytelling process**. The foundation of this process, as conceptualized by Lee et al. (2015) and expanded upon by Li et al. (2024), involves four stages: (1) analysis, (2) planning, (3) implementation, and (4) communication. (1) Analysis involves working with large, heterogeneous datasets from various platforms, identifying patterns and trends aligned with strategic goals using AI-powered tools. (2) Planning structures insights into coherent narratives, determining relevant findings and communication methods (Segel and Heer, 2010). (3) Implementation creates visual and textual narratives, often using tools like Tableau or Power BI (Chevalier et al., 2018). (4) Communication ensures tailored delivery for diverse audiences through dashboards, reports, or live presentations (Brehmer and Kosara, 2021). Collaboration in this context requires seamless transitions between stakeholders. Crisan et al. (2020) describe this as a "baton-passing" process, where roles such as data engineers, scientists, analysts, and decision-makers each contribute specific expertise (Gunklach et al., 2025). Data engineers ensure data readiness, data scientists uncover patterns, analysts contextualize findings, and decision-makers translate insights into actionable strategies. Domain experts play a critical role, offering deep knowledge of specific business domains to ensure that data stories are accurate and contextually meaningful (Hu et al., 2018). However, the necessity for multiple feedback loops between analysts and domain experts often creates bottlenecks (Lennerholt et al., 2021). For instance, in one of our interviews (DE23), a product manager described how weeks of back-and-forth with analysts were required to align KPIs with customer segmentation data, ultimately delaying critical business decisions.

A second challenge relates to the **role of data storytelling in enterprises** which is closely tied to supporting decision-making at operational, tactical, and strategic levels. At the operational level, storytelling helps monitor daily activities, such as tracking sales performance or optimizing production processes. Tactical storytelling informs medium-term decisions, such as marketing campaign adjustments or supply chain improvements. At the strategic level, narratives shape long-term objectives, like market expansion or digital transformation. Tory et al. (2022) identify two fundamental approaches to leveraging data for decision-making: "conversations with data" and "conversations around data". Conversations with data focus on straightforward tasks, such as checking KPIs or historical trends, requiring minimal interpretation. In contrast, conversations around data involve abstract, complex problems requiring synthesis, interpretation, and domain expertise to foster foresight and innovation. Achieving these decision-making goals requires more than just presenting accurate data; it also involves carefully navigating how stories are crafted and perceived. Baumer et al. (2022) emphasize that the framing and representation of data in storytelling are inherently political acts, shaped by organizational priorities and power dynamics. Choices about which data to highlight, how to present it, and whose perspectives to include can significantly influence decision-making outcomes. These decisions may unintentionally reinforce existing power structures or privilege certain viewpoints, potentially misaligning narratives with organization's broader goals.

Finally, the **domain-specific nature of data storytelling in enterprises** poses significant challenges, reflecting the diversity of industries and their unique data environments (Gunklach et al., 2023). Different business domains necessitate tailored approaches for contextualized data storytelling due to their distinct processes, priorities, and data characteristics (Kim et al., 2024). Enterprises often manage vast datasets sourced from transactional systems, IoT devices, and customer platforms, which are typically heterogeneous, encompassing structured, semi-structured, and unstructured formats (Kandel et al., 2012). Storytelling tools must not only harmonize these datasets at scale but also accommodate industry-specific requirements. For instance, healthcare organizations emphasize patient outcomes and regulatory compliance, while retail enterprises prioritize customer segmentation and inventory management. Customization of metrics, visualizations, and narrative elements ensures that storytelling aligns with industry standards, enhancing its relevance and impact.

2.2. Related work

Data storytelling has emerged as a pivotal practice in data communication, enabling users to derive actionable insights from complex datasets. Over the years, a wide range of tools has been developed to support the creation of data stories, each tailored to specific purposes and user groups. These tools vary significantly in their goals, target users, and features, often categorized by their focus on distinct stages of the storytelling process—analysis, planning, implementation, or communication (Chen et al., 2023). In addition to these functional distinctions, Li et al. (2024) suggest that data storytelling tools should also be evaluated by their collaboration dynamics, which describe how human and AI agents interact during the storytelling process. Collaboration types include creators, assistants, optimizers, and reviewers, with roles that may be performed either by humans or AI. For this review, we focus on data storytelling tools (see Table 1 for an overview) identified in recent literature (Chen et al., 2023; Li et al., 2024; Ren et al., 2023) that are suitable for enterprise use, producing outputs such as data stories, dashboards, or dynamic reports while supporting non-technical users in crafting meaningful narratives.

The **first group** of tools focuses exclusively on *implementation*. ChartAccent (Ren et al., 2017) enhances static charts with annotations, allowing users to emphasize specific data points or trends, while Text-to-Viz (Cui et al., 2019) automates the creation of infographics from textual descriptions. However, these tools lack integration with earlier stages like analysis and planning, limiting their utility in enterprise contexts where seamless transitions are essential. The **second group**, which spans *analysis to implementation*, includes tools that provide broader support for storytelling workflows. Erato Sun et al. (2022) integrates collaborative workflows, blending human creativity with AI assistance to structure narratives and generate visualizations. Notable (Li et al., 2023) employs embedding-based approaches to recommend narrative templates and supports dashboard creation. Similarly, Datashot (Wang et al., 2020) applies text mining and AI-driven layout optimization to generate fact sheets summarizing dataset trends. Calliope (Shi et al., 2021) combines Bayesian modeling for multivariate analysis with Monte Carlo search trees for visualization optimization. While these tools streamline workflows and enable insight generation, their outputs are often generic and require significant adaptation to align with specific enterprise contexts. The **third group** focuses solely

Table 1
This table compares data storytelling tools across the stages of analysis, planning, implementation, and communication. Most data storytelling tools lack comprehensive support across all stages, often producing static outputs and failing to integrate domain knowledge or enterprise workflows.

Tool	Analysis	Planning	Implementation	Communication
Group 1: Implementation				
ChartAccent (Ren et al., 2017)	*Not Available.*	*Not Available.*	Enhances static charts with annotations.	*Not Available.*
Text-to-Viz (Cui et al., 2019)	*Not Available.*	*Not Available.*	Converts text into static infographic visualizations.	*Not Available.*
Group 2: Analysis to Implementation				
Erato (Sun et al., 2022)	Collaborative annotations for exploration.	Story goals and chart editing for teams.	Timeline-based collaborative narratives.	*Not Available.*
Notable (Li et al., 2023)	*Not Available.*	Narrative templates based on data.	Interactive dashboards with story elements.	*Not Available.*
Datashot (Wang et al., 2020)	Summary statistics and correlations.	Trend extraction for summary visualization.	Static fact sheets.	*Not Available.*
Calliope (Shi et al., 2021)	Multivariate analysis.	Dynamic chart suggestions based on user input.	Automated visualizations and short narratives.	*Not Available.*
Group 3: Communication				
SketchStory (Lee et al., 2013)	*Not Available.*	*Not Available.*	*Not Available.*	Live storytelling with digital sketches.
Group 4: End-to-End				
CLUE (Gratzl et al., 2016)	Manual feature selection.	Templates for domain-specific contexts.	Manual visualization creation.	Static report.
InsideInsights (Mathisen et al., 2019)	Interactive variable filtering.	User-defined narrative structures.	Exploratory analysis with minimal automation.	Text-based commentary integration.

on *communication*. SketchStory (Lee et al., 2013) facilitates live presentations using hand-drawn sketches on digital interfaces, offering an interactive approach to engage audiences. However, it lacks integration with earlier stages like analysis or planning and does not support real-time feedback or adaptability during presentations. Finally, the **fourth group** encompasses end-to-end tools that span the entire storytelling process, from analysis to communication. CLUE (Gratzl et al., 2016) integrates exploratory data analysis with predefined templates to structure narratives, while InsideInsights (Mathisen et al., 2019) provides user-driven workflows that support all storytelling stages. Despite their comprehensive approaches, these tools rely heavily on manual input and offer limited assistance for chart creation or narrative generation.

The recent wave of GenAI systems has profoundly influenced how data stories are authored, interpreted, and communicated (Li et al., 2025). Large language models and multimodal models have expanded the output space of storytelling tools, enabling narratives, summaries, and visualizations to be generated from simple natural language prompts. Tableau Pulse (Tableau, 2025) delivers contextual insights by generating natural-language summaries of key metrics, surfacing anomalies and trends. Microsoft Power BI integrates Copilot and Smart Narrative features, which suggest visualizations and automatically produce textual explanations of dashboards (Microsoft Corporation, 2025) Such tools demonstrate the potential of GenAI to accelerate analysis, increase accessibility, and reduce reliance on technical specialists. These advances have diversified human–AI collaboration patterns: whereas earlier systems mainly positioned AI as a low-value assistant, newer approaches increasingly support AI-creator and human-reviewer workflows, where AI drafts stories and visualizations that domain experts validate and refine (Li et al., 2025). This shift lowers entry barriers for non-technical users, while reinforcing the continued importance of human oversight to ensure contextual accuracy and interpretability.

However, while GenAI excels at automating repetitive tasks, it cannot replace the nuanced understanding and contextual knowledge that domain experts bring to storytelling (Hutchinson et al., 2024; He et al., 2024). Generic AI-generated narratives and visualizations often require significant adaptation to align with business-specific goals, limiting their effectiveness as standalone solutions (Jung et al., 2025; Holzinger et al., 2022). Research has shown that domain-specific AI systems outperform more general-purpose models, highlighting the value of embedding domain expertise into tool design (Logeshwaran et al., 2024). Consequently, despite the diversity of both academic and industry tools, significant gaps remain for enterprise data storytelling (see Table C.6). Most systems fail to effectively integrate domain knowledge, resulting in fluent but generic outputs that require extensive manual refinement (Li et al., 2024). Comprehensive support across all storytelling stages is rare—only a small subset of tools span implementation (see group 1 in Table 1), analysis, planning, implementation (group 2), and communication (group 3). In group 4, we have clustered end-to-end tools, such as CLUE (Gratzl et al., 2016) and InsideInsights (Mathisen et al., 2019), which nominally cover the full pipeline but rely heavily on manual input: users must select features, design templates, and construct narratives largely on their own. This not only limits scalability but also leads to outputs that are generic rather than enterprise-specific. Outputs are often static, constraining adaptability to dynamic audiences or real-time updates, and iterative feedback mechanisms for collaborative storytelling are frequently absent (Ren et al., 2023). Integration with enterprise systems represents another critical gap, as most tools operate as standalone solutions disconnected from organizational workflows.

StoryPoint directly addresses these limitations by spanning all storytelling stages - analysis, planning, implementation, and communication - while embedding domain-specific knowledge leveraging GenAI. In the analysis stage, it contextualizes datasets and integrates enterprise goals to help users frame stories effectively. In planning, it offers customizable templates and tailored narrative suggestions aligned with organizational needs. For implementation, StoryPoint dynamically refines visualizations and narratives through natural language inputs and GenAI-supported recommendations, enabling seamless adjustments. In the communication stage, it supports real-time interaction through filtering, querying, and audience feedback mechanisms, ensuring relevance and engagement. Its secure integration with enterprise systems ensures that insights remain aligned with organizational workflows, making it a comprehensive and adaptable solution for domain-specific data storytelling in enterprise contexts.

Table 2
This table summarizes semi-structured interviews with 12 domain experts, detailing their Python proficiency, dashboard utilization, and dashboard creation skills, measured on a 1–7 Likert scales. It also includes their business domains and interview durations, providing a comprehensive overview of participant expertise.

Interview	Python	Dashboard utilization	Dashboard creation	Business domain	Duration
DE01	3	6	4	Product Management	00:37
DE02	2	5	3	Product Management	00:46
DE03	3	5	2	Product Management	00:31
DE04	3	6	4	Controlling	00:28
DE05	3	5	3	Product Management	01:16
DE06	2	6	3	Product Management	00:40
DE07	4	7	3	Controlling	01:11
DE08	3	6	4	Controlling	01:07
DE09	3	5	2	Product Management	00:59
DE10	4	6	3	Product Management	00:39
DE11	2	5	3	Purchasing	00:37
DE12	3	5	2	Purchasing	01:01

3. Design of StoryPoint

In this section, we will explain how we designed and built *StoryPoint*. The basic user interaction concept of *StoryPoint* is illustrated in the graphical abstract.

3.1. Requirement elicitation

To build a literature-grounded and user-centered data story authoring tool, we first conducted twelve semi-structured user interviews including think-aloud sessions with domain experts to receive an understanding of their needs and requirements for a data story authoring tool. The interviewees were domain experts who were all potential users of the system due to limited skills in data visualization. Participation was voluntary and no compensation was offered, as participants acted within their professional capacities and participation time counted as hours worked. We provide information regarding the participants' role and business domain in Table 2. To capture technical experience, we used Python proficiency as a proxy, since Python is the most widely adopted language in enterprise analytics, offers extensive visualization libraries (e.g., Matplotlib, Seaborn, Plotly), and is frequently employed as a benchmark in HCI studies (Li et al., 2023; Lin et al., 2025). At the same time, to avoid over-reliance on a single measure, we also assessed dashboard utilization and dashboard creation experience, which together provide a broader view of participants' technical expertise. On a scale of 1 (worst) to 7 (best), the average Python skills of participants was 2.92, dashboard utilization 5.58, and dashboard creation 3.0. Participants' ages ranged from 28 to 45 years, with a mean age of 34 years. The interview guideline consisted of 29 questions and each interview lasted on average 49 min.

We structured the interviews into four parts: (1) general questions about the participant's role in the enterprise and daily tasks, (2) experience and problems with data visualization tools (e.g., Tableau, PowerBI), (3) requirements for a data story authoring tool, and (4) think-aloud-session with Tableau including its "AskData" (Setlur et al., 2016) feature (see appendix for interview questionnaire). For the think-aloud session, users were tasked to create a simple data story with a dataset of their choice and vocalize all thoughts loud. We conducted the think-aloud session to collect domain experts' problems when using Tableau and gather further requirements for a data story authoring tool. We chose Tableau because it promises easy-to-use, state-of-the-art data visualization (Ruoff et al., 2023) and is currently being used by the industry partner. We recorded all interviews on audio and transcribed them. Following Li et al. (2024) and Lee et al. (2015), we relied on an established set of four data storytelling process steps (analysis, planning, implementation, and communication) to carry out a selective coding of our interview transcripts. Selective coding is a stage in grounded theory research that serves to organize the analysis around a core set of variables (Bryant and Charmaz, 2007). Using a selective coding process allowed us to scaffold our analysis around what features domain experts need to create data stories. Some participants made explicit statements, for instance, *For story composition, I want to see a list of potential charts that I can rely on*. Following Li et al. (2024), we integrated that into the step implementation. Further, some statements were implicit. *I need a simple way to create visualizations*. To derive requirements, we grouped implicit and explicit statements into the data storytelling process steps (Li et al., 2024; Lee et al., 2015).

3.1.1. Analysis requirements

Analysis is fundamental to the data storytelling process, as it provides users with the insights needed to craft meaningful narratives. Participants emphasized the need for systems to visualize data quality and relationships, enabling quick assessments of dataset usability (**R1**). *"I need to understand the data's structure quickly to decide if it's usable for my goals"* (DE3). Another common need was accessible previews and summaries to provide an overview of datasets without requiring extensive manual inspection (**R2**). Participants also highlighted the importance of seamless integration with enterprise infrastructure to streamline workflows (**R3**). *"Integrating with our existing databases would streamline the process immensely"* (DE6). Data cleaning support emerged as a critical requirement, with participants requesting guidance on handling missing or anomalous values (**R4**). Furthermore, extracting metadata and contextual knowledge from datasets was highlighted as essential for framing the analysis (**R5**). *"Extract metadata to set the data into a context"* (DE2). Finally, participants emphasized the value of defining the story purpose at the start of the process to ensure alignment and focus (**R6**). *"Letting me input the story's goals upfront would make the process more focused"* (DE10).

3.1.2. Planning requirements

The planning phase involves structuring the narrative to guide the audience through insights effectively. Participants frequently requested customizable templates to save time and ensure consistency across stories (**R7**). *"Predefined templates would save time and ensure consistency across reports"* (DE2). Multi-page story creation with progression recommendations was also seen as vital (**R8**). *"For me, it's important to build a narrative that flows from high-level insights to detailed analysis"* (DE10). Participants sought systems that could suggest layouts and visual styles tailored to specific datasets and goals (**R9**). *"Sometimes I'm not sure what layout works best"* (DE11). Additionally, pre-set themes tailored to branding and domain-specific requirements were requested to enhance consistency and professionalism (**R10**). *"I usually create data stories for several business domains that have a different structure"* (DE12).

3.1.3. Implementation requirements

The implementation phase requires tools to create clear and engaging narratives that connect visualizations to insights. Participants emphasized the importance of systems recommending suitable charts and narrative elements to streamline the storytelling process (**R11**).

Table 3
This table outlines the derived design principles for developing a data story authoring tool for enterprises. These principles include integrating domain expertise, intelligent visualization creation, structured layouts, contextual explanations, advanced presentation options, and continuous feedback.

Design principle	Description	Supported by literature	Influenced by requirements
Integration of Domain Expertise	Support domain experts in all steps of the enterprise data storytelling process, by dynamically combining their contextual inputs—such as metadata and story objectives—with GenAI capabilities, ensuring narratives and visuals align with organizational goals and are tailored to enterprise needs.	Lee et al. (2015), Gunklach et al. (2023), Ramm et al. (2021), Kim et al. (2024), Kandel et al. (2012)	R4, R5, R6, R7, R17
Intelligent Visualization Creation	Provide domain experts with guidance in the form of automatically generated charts, enabling them to select and adjust visuals using natural language input to refine storytelling objectives.	Shi et al. (2021), Wu et al. (2022), Ye et al. (2024), He et al. (2024)	R1, R2, R3, R12, R13
Structured Data Story Layouts	Allow domain experts to select predefined data story layouts that promote narrative coherence, contributing to uniform and easily navigable data stories.	Liem et al. (2020), Martinez-Maldonado et al. (2020), Shao et al. (2024)	R7, R8, R9, R10, R11
Contextual Explanations	Provide domain experts with options to include explanations such as overarching story purposes and interpretative elements, including the meaning of charts and their KPIs, to contextualize insights.	Masson et al. (2023), Li et al. (2023), Microsoft (2024)	R6, R7, R10, R14, R16, R20
Advanced Presentation Options	Include advanced presentation options, such as AskData features and full-screen modes, to support domain experts in answering audience questions and enhancing engagement during presentations.	Tory et al. (2022), Sarikaya et al. (2019), Tableau (2024)	R17, R18, R19
Continuous Feedback	Enable audience members to give feedback directly on specific aspects of the data story, facilitating iterative improvement and effective collaboration.	Elias et al. (2013), Hu et al. (2018), Haug et al. (2023)	R20, R21

"The system could suggest charts that best match my data and purpose" (DE7). Enabling visualization adjustments through natural language input was another key request (**R12**). *"It would save so much time if I could adjust visuals just by typing commands"* (DE6). Participants frequently highlighted the need for dynamic integration of KPIs and calculations into narratives to align insights with organizational goals (**R13**). *"I would need to include key metrics dynamically within the narrative"* (DE9). Tools to emphasize key insights using design elements such as colors and fonts were also seen as essential for drawing attention to critical data points (**R14**). *"The ability to emphasize certain insights would make the story more engaging"* (DE3). Additionally, participants sought systems that dynamically adapt narratives based on contextual information, including audience-specific phrasing (**R15**). *"If the charts and text could be quickly adapted to different audiences, that would be nice"* (DE6). Lastly, generating a concise page purpose to connect charts and narratives cohesively was seen as essential for maintaining structure (**R16**). *"I'd like to have a concise description of each page"* (DE11).

3.1.4. Communication requirements

The communication phase focuses on delivering the story to the audience effectively, with tools that support customization, interaction, and feedback. Participants requested dynamic filtering of story elements for different audience types to ensure relevance (**R17**). *"I need to adjust what's shown depending on who I'm presenting to"* (DE3). Additionally, participants requested options to export stories in multiple formats to accommodate various sharing needs (**R18**). *"Having different export formats would make the tool more versatile"* (DE12). Interactive querying during presentations was frequently mentioned as essential for addressing audience questions dynamically (**R19**). *"I want to be able to answer questions on the spot during presentations"* (DE9). Finally, participants highlighted the importance of audience feedback mechanisms (**R20**) and revision tracking for iterative improvement (**R21**). *"I struggle to get structured feedback on my stories. It would help me if people could give direct feedback on certain elements"* (DE12). *"A log of changes would help track the evolution of the story"* (DE4).

3.1.5. Design principles

The design principles for StoryPoint were developed through a systematic process grounded in human–computer interaction research relying on requirements from interviews and literature (Norman, 1983; Chao, 2009). For the rigorous literature-grounded approach, we followed the method of Webster and Watson (2002) to conduct a systematic literature review to derive a set of requirements for the design of a data story authoring tool for enterprises from existing literature. The search string was developed iteratively, starting with exploratory searches using terms like "data storytelling", "data stories", and "narrative visualization". Based on this, we identified additional relevant terms, such as "data journalism" and synonyms like "narrative visualization", particularly prevalent in seminal works (Segel and Heer, 2010; Hullman and Diakopoulos, 2011) The final search string incorporated "storytelling" and "narratives" in combination with "data visualization", "information visualization", "visual analytics", and "business intelligence" to ensure comprehensive coverage (see appendix). Using this search string, we retrieved 1507 unique articles from IEEE Xplore, ACM DL, and Scopus after deduplication. In the selection process, we manually excluded 1281 articles by carefully scanning the title-, abstract-, and keyword-section and by applying the following selection criteria: We included articles that investigated the implementation and evaluation of data stories. And we excluded articles that relied on data storytelling techniques in virtual or augmented reality, videos, or comics as these media formats are not frequently used in enterprise data storytelling. Following the same criteria for a full-text review, 71 articles remained. Lastly, we performed a forward and backward search and included another 21 studies. This process yielded 92 articles for intensive analysis.

Using a selective coding approach, we categorized requirements and identified higher-level themes—design principles—that addressed recurring user needs and underlying motivations. The process began with clustering related requirements to uncover broader themes. Insights from the systematic literature review, such as methods, techniques and challenges, were analyzed alongside user requirements derived from interviews. For instance, studies emphasizing the importance of

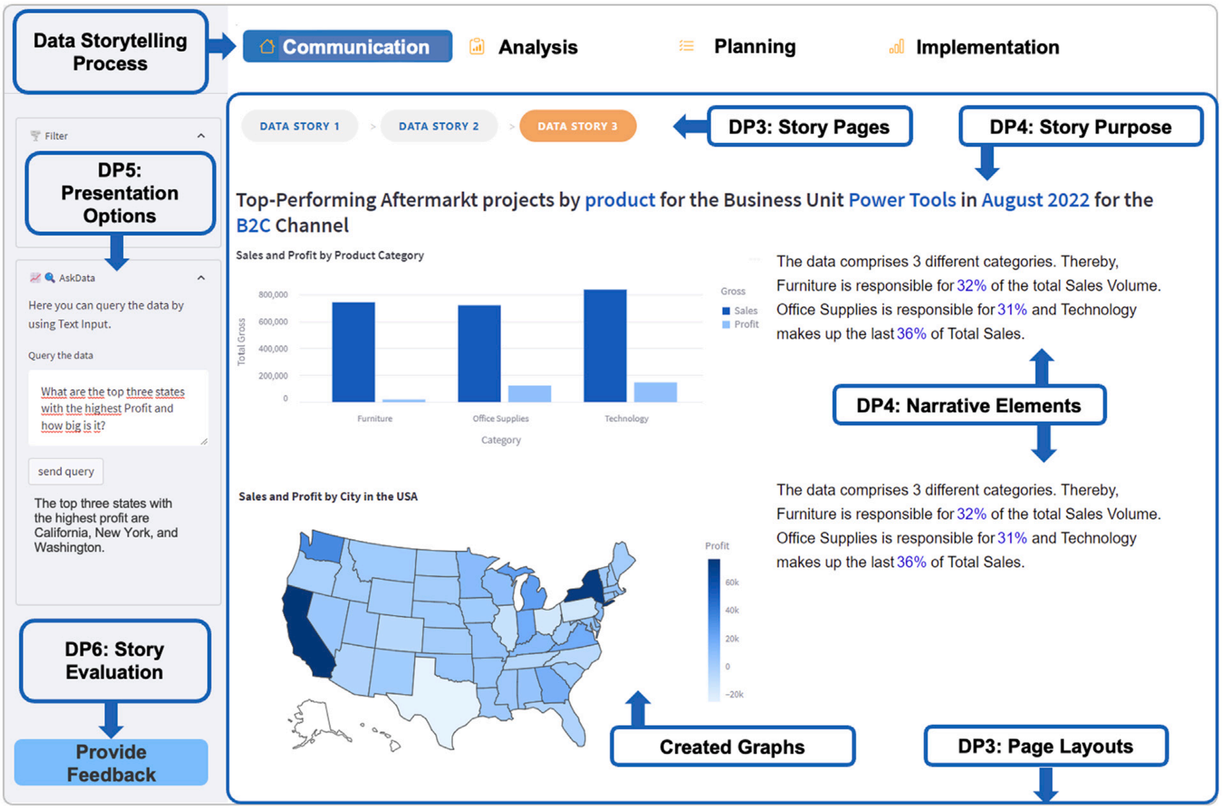

Fig. 1. This figure shows an example data story created with *StoryPoint*, highlighting sales data across product categories and regions using visualizations and narrative elements. The top includes a structured narrative with multiple pages, while the sidebar allows interactive exploration of sales insights through natural language queries, supported by visualizations and feedback options.

context in narrative visualizations (Gunklach et al., 2023; Kim et al., 2024) informed requirements like extracting contextual metadata (R5) and adapting narratives to specific contexts (R15). These themes were then translated into actionable principles, ensuring the system provided practical functionalities to meet identified challenges. To maintain comprehensive coverage, a traceability matrix (see Table 3) mapped requirements to their respective principles. This structured approach ensured that the principles were both user-centered and empirically grounded.

3.2. StoryPoint

StoryPoint is built as a responsive web-based application, explicitly grounded in the six design principles presented (see Table 3). Each principle is instantiated in the tool's usage flow, which sequentially leads domain experts through the four stages of enterprise data storytelling: analysis, planning, implementation, and communication. For example, *StoryPoint* integrates domain expertise (**DP1**) by combining metadata and story purposes with GenAI capabilities, supports intelligent visualization creation (**DP2**) through natural language and example-based chart generation, and ensures narrative coherence with structured layouts (**DP3**). Contextual explanations (**DP4**) enrich visualizations with interpretative meaning, while advanced presentation options (**DP5**) such as AskData enhance audience engagement. Finally, continuous feedback mechanisms (**DP6**) allow iterative refinement of the story. Each page includes a sidebar that offers additional settings. Fig. 1 shows an example data story created with *StoryPoint*. Together, these features not only guide users step by step through the authoring process on any device but also enable us to empirically evaluate how well the design principles support domain experts in creating enterprise data stories. In the following, we describe *StoryPoint* and its features by relying on an example enterprise use case that illustrates how domain experts could create data stories with *StoryPoint*.

Sarah, a newly onboarded consultant at a boutique consulting firm, receives sales data from the client's Chief Sales Officer and is tasked to create an initial report, containing the first important insights regarding sales until end of day. However, Sarah acknowledges her limited proficiency in data visualization techniques and utilizes *StoryPoint* to visualize the data. For **analysis**, she uploads her data and *StoryPoint* automatically generates an exploratory data analysis report. She goes through the report and uncovers key insights of her dataset, such as how various columns are distributed and correlated. At the end of the analysis phase, Sarah is prompted to input additional contextual information, such as the intended purpose of the story or specific visualizations she wants to include. For example, she specifies that the data story should focus on identifying top-performing product categories. These inputs are integrated into StoryPoint's LLMs, ensuring that the charts and narratives align closely with Sarah's analytical objectives and storytelling requirements. In the **planning** stage, she relies on these insights to create the first page of the data story. For that, she chose one of the page layouts that are provided by *StoryPoint* (**DP3**). If none of the page layouts fit her page, she could create a custom page layout. Sarah's chosen layout accommodates two charts, vertically arranged, accompanied by narrative elements on the right-hand side. Next, in the **implementation** phase, *StoryPoint* guides her through the creation of the charts and narrative elements of the first page. For chart creation, the tool offers two distinct modes (**DP2**): (1) natural language input, where she can articulate her requirements conversationally, or (2) example visualizations that serve as a springboard for her creativity (see Fig. 2). Sarah's creative process unfolds as she selects the second option for her first chart. She peruses a curated list of example visualizations and chooses one that vividly illustrates the total sales amount over time. Next, she refines the chart's appearance using the

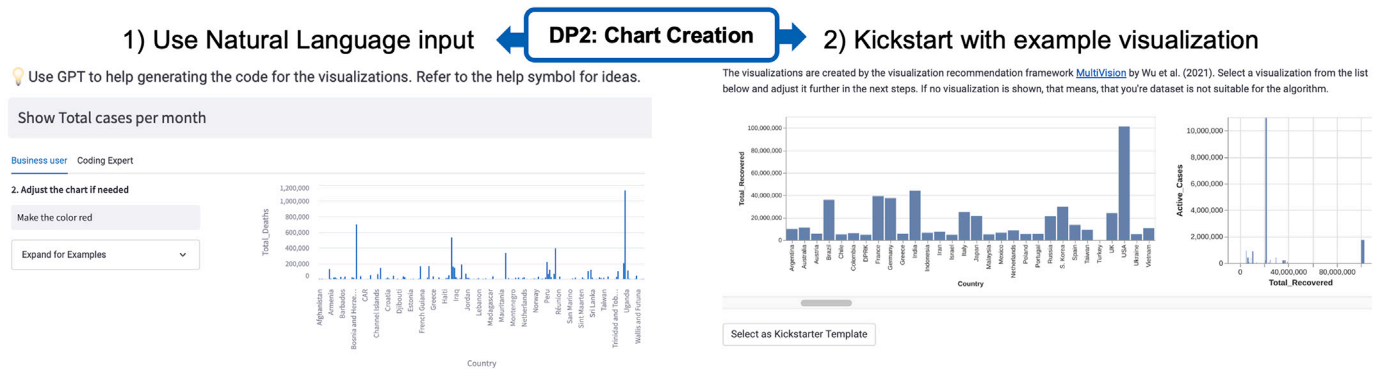

Fig. 2. This figure illustrates chart creation options of *StoryPoint*. On the left, users can generate visualizations using natural language input, with options to adjust and refine the chart. On the right, users can kickstart their data exploration with example visualizations generated by a recommendation framework, providing a quick starting point for customization.

natural language input *Add the dimension category to the chart and give each category a different color*. After careful inspection and iterative refinements, she achieves the desired result—an interconnected line chart that resonates with her analytical intentions. *StoryPoint* also equips her with various explanatory options for the chart (**DP4**). She opts for the description: *This chart portrays the annual sales figures for three distinct categories: furniture, office supplies, and technology, spanning the years 2014 to 2017*. She augments the text with insights garnered during her exploration phase, adding a paragraph about the trends she observed in these categories from the exploratory data analysis report. Then she continues with the second chart. Her simple prompt, *Show the top 10 best-selling products per category as a bar chart* swiftly produces a chart that aligns with her expectations (**DP3**). She opts for *StoryPoint*'s pre-generated narratives (**DP4**). To enhance the narrative's impact, she inserts a dynamic variable calculating the revenue, allowing for automatic updates. After creating both charts and descriptions, *StoryPoint* automatically creates potential story purposes (**DP4**) that capture the content of the page. She selects the one that encapsulates her analytical objectives: *Analyzing Best Sellers and Sales Evolution: Insights from Top 10 Products and Category Performance over Time*. While the option to expand her narrative with additional pages exists, Sarah defers that decision for later. She concludes her work by saving the crafted data story and sending a link to her client. When accessing the data story (**communication** phase), the client is presented with the opportunity to engage with the data through two distinct avenues: either by perusing the data story narrative or by utilizing *StoryPoint*'s Ask Data feature to pose natural language questions (**DP5**). Additionally, the client can opt to provide direct feedback to specific elements within the data story by simply clicking on them. This action triggers the appearance of a menu within the sidebar, offering the client the ability to assess the element using a star rating system or to furnish feedback in written form (**DP6**).

3.3. StoryPoint features

The architecture of *StoryPoint* is designed to facilitate seamless human-AI interaction for data storytelling. In line with the 5Vs of big data (Anuradha et al., 2015), it scales to large datasets (volume), supports timely and interactive updates (velocity), accommodates heterogeneous data formats (variety), grounds generated narratives in source data to ensure reliability (veracity), and embeds domain-specific goals to keep insights actionable and relevant (value). Built on a web-based client–server model, it integrates technologies such as the Streamlit Python library for the web client, React for front-end components, and Python for back-end processing. Key elements include data storage in CSV format and the use of VegaLite for generating visualizations (Satyanarayan et al., 2016). The system comprises four main features: (1) visualization creator, (2) chart description editor, (3) story purpose editor, and (4) data story viewer (see Fig. 3).

StoryPoint's innovation lies in its ability to dynamically adapt its LLMs to domain-specific contexts, such as sales or logistics, by leveraging metadata extracted from the dataset. During the analysis phase, domain experts can enhance this contextualization by inputting their story purpose and specific requirements, such as key visualizations or narrative goals. This combination of metadata and user-defined inputs ensures that the outputs - charts, narratives, and story purposes, are purpose-aligned and data-driven (**DP1**). The tool features a **visualization creation** feature, leveraging GPT-3.5 Turbo and GPT-4 for dynamic chart generation (**DP2**). A deterministic model configuration minimizes errors, and any issues prompt user re-engagement. Additionally, the chart recommendation network from MultiVision (Wu et al., 2022) suggests VegaLite specifications, enhancing user-driven creation. Confirmed visualizations trigger a GPT-4 instance to generate concise **chart descriptions** (**DP4**). These descriptions are based on chart features and enhanced dataset samples, providing user-centered explanations in bullet-point format. A custom text editor allows users to refine these descriptions and to integrate dynamic variables. For that purpose, we developed a custom streamlit component that allows users to integrate dynamic variables into the editor and published it as a python package. The **story and page purpose editor** (**DP4**) utilizes GPT-4 to create cohesive dashboard titles, summaries, and page-specific names, which users can adjust to refine their narratives. The final iteration of the **data story viewer** incorporates three interactive features (**DP5**): (1) the ask data agent, (2) interactive filters, and (3) a feedback mechanism. The ask data agent uses GPT-4 to interpret user queries, executing Pandas queries directly on the dataset. Interactive filters enable users to customize numeric and nominal feature views, with options to clear filters. The feedback mechanism (**DP6**), inspired by Haug et al. (2023), allows users to provide feedback on specific story elements by selecting a category, assigning a rating, and adding comments. Feedback is stored systematically with screen coordinates for iterative improvements.

4. Formative evaluation

Throughout the development phase, our approach involved conducting both formative and summative evaluations to assess how effectively *StoryPoint* facilitates the creation of data stories for domain experts. The formative evaluation provided insights into the tool's ongoing development, ensuring that it aligned with user needs and expectations. In contrast, the summative evaluation, which involved real-world data visualization use cases, was conducted to comprehensively assess the system's usability and performance. Table 4 presents a comprehensive summary of all participants who took part in the formative evaluation. We recruited eight domain experts aged 21–52 (mean = 30) with diverse backgrounds and varying levels of expertise. We assessed participants' proficiency in Python, dashboard utilizing,

Table 4
This table summarizes the formative evaluation with eight domain experts, detailing their self-reported Python proficiency, dashboard skills, and business domains. Participants showed strong dashboard utilization skills (mean = 4.5) but moderate Python (mean = 3.3) and dashboard creation (mean = 3) skills.

Interview	Python	Dashboard utilization	Dashboard creation	Business domain	Duration
DE13	2	4	3	Purchasing	00:20
DE14	4	4	2	Purchasing	00:30
DE15	4	5	3	Logistics	01:00
DE16	4	7	5	Consulting	00:25
DE17	2	3	2	Consulting	00:40
DE18	2	4	3	Manufacturing	00:30
DE19	5	7	6	Information Technology	00:50
DE20	5	4	3	Research	00:40

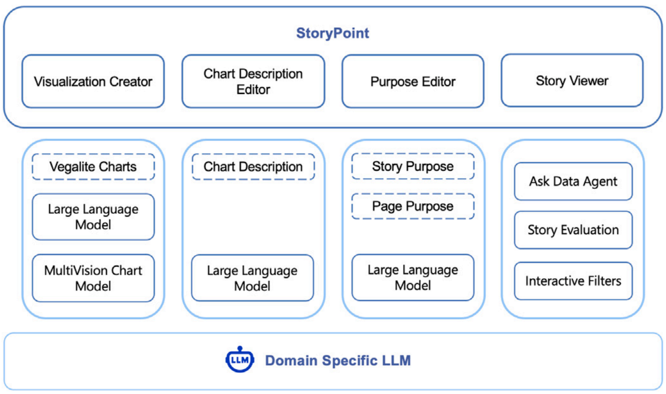

Fig. 3. This figure demonstrates the architecture of *StoryPoint*. The system integrates components like the visualization creator, chart description editor, purpose editor, and story viewer to support the end-to-end workflow of data storytelling.

and dashboard creating skills using a Likert scale ranging from 1 (worst) to 7 (best). On average, the participants demonstrated a low proficiency in Python, with a mean rating of 3.5. When it came to utilizing dashboards, the participants exhibited a strong aptitude, garnering a mean rating of 4.75. Lastly, in creating dashboards, their skills averaged 3.3.

4.1. Method

Formative evaluations aim at collecting information to improve an artifact (Ritchie et al., 2013). We, therefore, conducted think-aloud sessions with eight domain experts to evaluate how *StoryPoint*'s design supports them in creating data stories. For the think-aloud sessions, we tasked the participants to choose a dataset and subsequently explore the data, and create a data story. In our formative evaluation, we aimed at evaluating the functionality and usability of our tool, to improve it in preparation for our final, summative evaluation.

4.2. Findings

First of all, domain experts expressed a high satisfaction with the **ease of use** of *StoryPoint*. As one user remarked, *"I really liked the prescriptive nature of the natural language prompts. Instead of saying what I want on the x- or y-axis, I was able to just say what the outcome should be"* (DE14). This approach made the tool accessible even for domain experts without prior experience in data visualization. *"Compared to Excel, the tool is really easy to use for the first time and I don't feel that I have to watch tutorials to create my first charts"* (DE20). They emphasized that this simplicity inspired confidence in them. *"After just creating one data story, I now feel confident enough to be able to use the tool self-reliantly"* (DE17). Further, domain experts liked the **usefulness** of *StoryPoint* for their data visualization needs. Some expressed a willingness to integrate it into their daily workflow for rapid data story creation. *"I would definitely give the tool a try for my management presentations if you make a public productive version"* (DE16). Engaging with the visualization features has also been a source of **enjoyment** for users, with one user stating, *"The Chart is fine like this but I want to continue and test what else the model is capable of creating"* (DE15). Another user highlighted the **learning** experience, he mentioned that he *"was positively surprised by the capabilities of the tool. Just by playing around with the visualization feature I learned a lot about the data"* (DE18). Users commend the modern and customizable look and feel of the **charts** generated, and the *"options for configuration"* (DE17). Feedback regarding the **narrative elements** was generally positive. Users liked the option to describe the graph using text and dynamic variables. "This suggests room for improvement in tailoring the narrative elements to better align with specific user contexts and preferences" (DE16). Users were satisfied with the system's **response time**, considering it suitable for a web-based application. One domain expert concluded: *"Good response time for a web-based AI application"* (DE14). Users highlighted the efficiency of *StoryPoint*. "This is the first data story that I developed, and yet it only took 20 min from the data exploration until the finished story" (DE18). This feedback underscores the system's effectiveness in streamlining the data storytelling process. Users recognized the trade-off between customizability and simplicity when **creating data stories**. Novice users tend to prioritize simplicity, while others desire more customization options. Importantly, users appreciate the dynamic updating of charts when applying filters, indicating a preference for interactive and responsive features. However, domain experts expressed a need for the ability to verify the results generated by the LLM, especially for business use cases. *"Before employing this tool in a business use case, I need to be sure that the data in the charts is correct. At the moment I am skeptical about the consistency of the output of large language models, so I would need the option to verify the results to gain trust in the visualizations"* (DE19). In addition to that, they expected **consistency** in the charts generated, *"when the tool doesn't provide the expected output, one can become frustrated"* (DE15).

4.3. Additional requirements

During the formative evaluation of *StoryPoint*, participants provided invaluable insights, accompanied by specific requirements to enhance their interaction with the tool. Two pivotal requirements emerged from this process, which guided our improvements (see Fig. 4 for the implementation of these features). Firstly, one user remarked, *"I think the tool in itself is really powerful, but someone like me probably doesn't know how to utilize its full potential, so I would suggest some example prompts"* (DE16). In response, we introduced an expandable widget within *StoryPoint*, filled with example prompts to provide the desired guidance for natural language prompts (**R22**). Secondly, we addressed the need for automatically generated narrative elements (based on user-defined contexts and visualizations) to eliminate the intimidation of a blank canvas when composing chart descriptions (**R23**). One user stated *"It would be good to have automatically generated narrative*

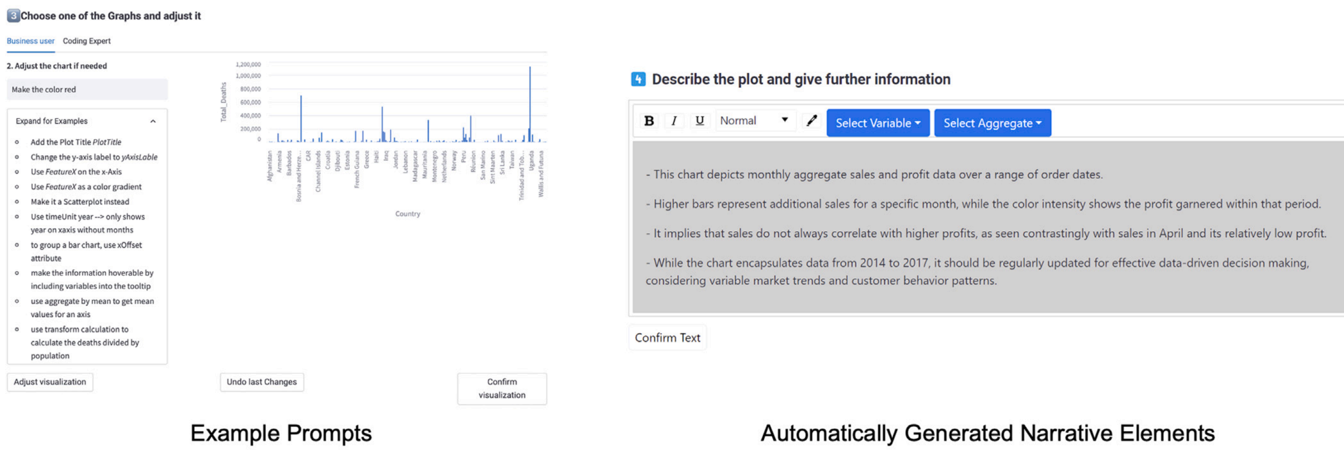

Fig. 4. This figure underscores the main features implemented in *StoryPoint* after the formative evaluation. On the left, users can adjust visualizations with example prompts and editing options. On the right, automatically generated narrative elements provide descriptive insights and contextual information for the charts, supporting users in crafting effective data stories.

elements that I can rely on when writing the chart descriptions. The editor would not be blank" (DE17). Additionally, we've also attended to several minor requirements, such as fullscreen option for the data story viewer (**R24**). Furthermore, in response to user feedback, we implemented two additional features to enhance *StoryPoint*'s functionality and address potential challenges. Firstly, we incorporated the ability for users to make adjustments to data stories after their initial creation (**R25**), allowing for flexibility in an enterprise setting where timelines and bandwidth constraints may necessitate adjustments post-creation (Chen et al., 2023). Secondly, beneath the text field containing generated narratives, we added a warning/hint emphasizing that narratives generated by the LLMs may not be fully conclusive (**R26**), encouraging users to validate content for accuracy (van der Lee et al., 2021). These enhancements contribute to a more robust and user-friendly data storytelling tool, aligning with the evolving needs and challenges of enterprise settings.

5. Summative evaluation

In contrast to the formative evaluation, a summative evaluation of an artifact is concerned with its impact and resulting outcomes (Ritchie et al., 2013). Following this dual evaluation approach ensured that *StoryPoint* not only met immediate requirements but also excelled with regards to its impact and outcomes in practical, real-world applications. Building on Shao et al. (2024) and Li et al. (2024), we structure our summative evaluation around three guiding questions: (1) Can the tool guide domain experts in creating efficient and effective data stories? (2) How is the quality of the created data stories? and (3) To what extent can the tool be applied for real-world data storytelling use cases? To address these questions, we conducted three dedicated summative evaluations. In the first summative evaluation, we compared *StoryPoint* with a benchmark system on efficiency, workload, and usability, considering both task completion times and subjective measures such as NASA-TLX and SUS. In the second summative evaluation, we conducted a large-scale study with 104 crowdworkers from Prolific, who evaluated the data stories created in the first summative evaluation from both systems along dimensions such as fluency, clarity, and trustworthiness. While this large-scale study provided robust comparative ratings, we acknowledge that crowdworkers may not fully capture the contextual nuances of enterprise data storytelling (Huang et al., 2023). To address this, we complemented it with a third summative evaluation, where domain experts from three different enterprise domains each applied *StoryPoint* to a distinct data visualization use case. Together, these three studies provide a comprehensive assessment of *StoryPoint*, demonstrating its efficiency and output quality in controlled evaluations as well as its applicability and usefulness in real-world practice.

5.1. Evaluation of StoryPoint's ability to create data stories

In this evaluation, we aimed to compare the efficiency, workload, and usability of *StoryPoint* in creating data stories with that of a benchmark system (see Fig. 5). We recruited eight additional domain experts aged 22–58 (mean = 29) with diverse backgrounds and varying levels of expertise. We assessed participants' proficiency in Python, dashboard utilizing, and dashboard creating skills using a Likert scale ranging from 1 to 7. On average, the participants demonstrated a low proficiency in Python (M = 2.9), utilizing dashboards (M = 3.75), and creating dashboards (M = 2.625). Each participant spent 34 min on average completing the study.

5.1.1. Method

The evaluation was structured as a small-scale experimental study employing a 2 × 2 design, encompassing four distinct groups denoted as follows: Group 1 engaged with *StoryPoint* initially for the sales task, followed by the benchmark for the subsequent customer behavior task. Conversely, Group 2 commenced their tasks with *StoryPoint* focused on customer behavior, transitioning to the benchmark for the subsequent sales task. Group 3 began with the benchmark for the sales task, subsequently utilizing *StoryPoint* for the ensuing customer behavior task. Finally, Group 4 initiated with the benchmark for the customer behavior task, followed by *StoryPoint* for the sales task. This experimental design allowed for a comprehensive exploration of both systems (*StoryPoint* and benchmark) across two distinct tasks—sales data and customer behavior data. In creating our benchmark, we sought to align it with Calliope (Shi et al., 2021), a recognized system for generating data stories. However, due to data privacy regulations imposed by our industry partner, we were unable to directly utilize Calliope. Therefore, we reengineered a similar benchmark by incorporating a Monte Carlo search tree for automated chart generation (Shi et al., 2021) and integrating functionalities from Voyager2 (Wongsuphasawat et al., 2016) to refine the charts. For narrative development, we generated concise text summaries of the charts similar to Calliope (Shi et al., 2021). The benchmark was implemented by one of the authors of this paper and

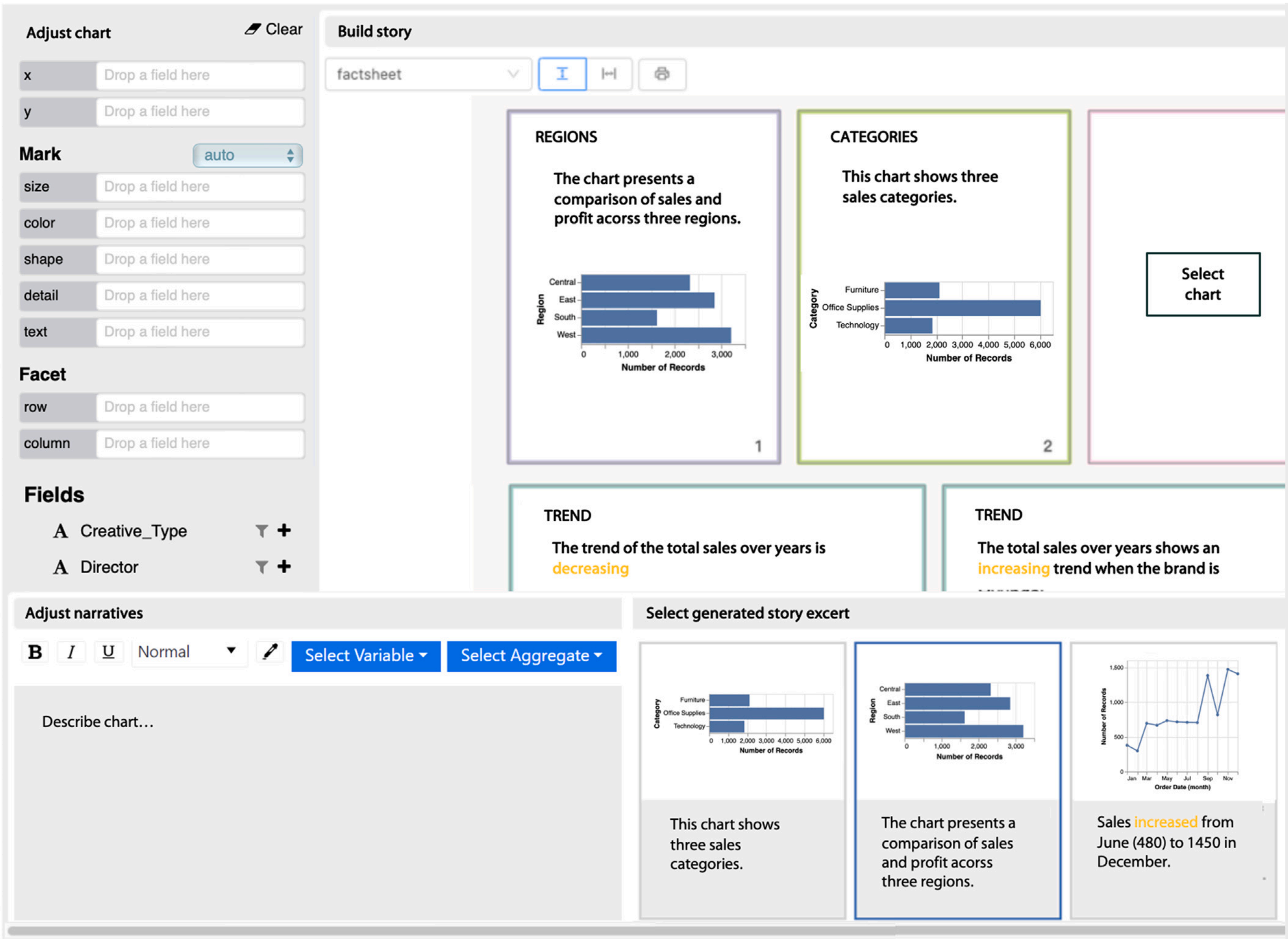

Fig. 5. This figure shows the user interface of the benchmark tool used in Section 5.1. The interface allows users to adjust charts by modifying axes, marks, and fields on the left panel. The central area supports building data stories with pre-defined templates, where charts and associated narratives can be arranged. At the bottom, users can refine automatically generated narratives, select story excerpts, and provide contextual descriptions to create a cohesive and customized data story.

subsequently evaluated by the rest of the author team to ensure comparability with Calliope. For the sales data task, the requirements were a bar chart displaying total sales for each product category and a bubble chart illustrating the relationship between profit, sales region, and customer data. Likewise, for the customer data task, participants were asked to create a bar chart showcasing customer satisfaction across various age groups and a donut chart visualizing factors contributing to cart abandonment. Following the chart creation, participants were required to provide detailed descriptions for each chart and conclude the task by composing a title along with a comprehensive description of the entire data story.

After completing a demographic questionnaire, participants were informed that they would *"use two different systems to create two different data stories"*. For each treatment, participants went through a tutorial in which they watched a video and then practiced what they just saw, by creating a simple data story with the open source movie dataset from IMDb. There were 5 steps in the tutorial: (1) select and upload the movie data set, (2) create a chart showing the number of movies per year, (3) create a table showing the movies with IMDb rating 9.2, (4) provide detailed descriptions for each chart and (5) conclude the task by composing a short summary for both charts. Following the tutorial, participants performed the relevant task using the corresponding system aligned with the appropriate treatment. After each treatment, participants evaluated the system using the NASA Task Load Index (Hart, 2006) scale (NASA-TLX), and the System Usability Scale (SUS) (Brooke, 1996). These metrics aimed to gauge the participants' usability and workload associated with each system.

5.1.2. Results

We employed statistical tests to analyze the results, utilizing Student's t-test for normally distributed values (e.g., time) and the Wilcoxon Signed Rank test for non-normally distributed values (e.g., workload, usability, and user experience) through the stats module from scipy (Virtanen et al., 2020). The independent two-sample t-test identified a significant effect of the condition on **time** ($p = 0.000$, $t = -6.84$). Participants spent a significantly lower amount of time on the task to create the data story with *StoryPoint* (M = 4 min 31 s, SD = 58 s) compared to the benchmark (M = 10 min 11 s, SD = 2 min 8 s). One domain expert expressed after the experiment, "*It was feasible to develop the data story using the benchmark system, but I simply needed more time, especially for fine-tuning the narratives and charts*". Regarding **workload** (NASA-TLX), The Wilcoxon Signed Rank test revealed significant effects of the condition on mental demand ($p = 0.0078$), performance ($p = 0.0156$), effort ($p = 0.0078$), and frustration ($p = 0.0269$). The participants with *StoryPoint* rated their performance significantly higher than with the benchmark, while the mental demands, effort, frustration, and time demands were lower. However, there was no significant effect on physical demand ($p = 0.0707$). The **usability** (SUS) results indicated

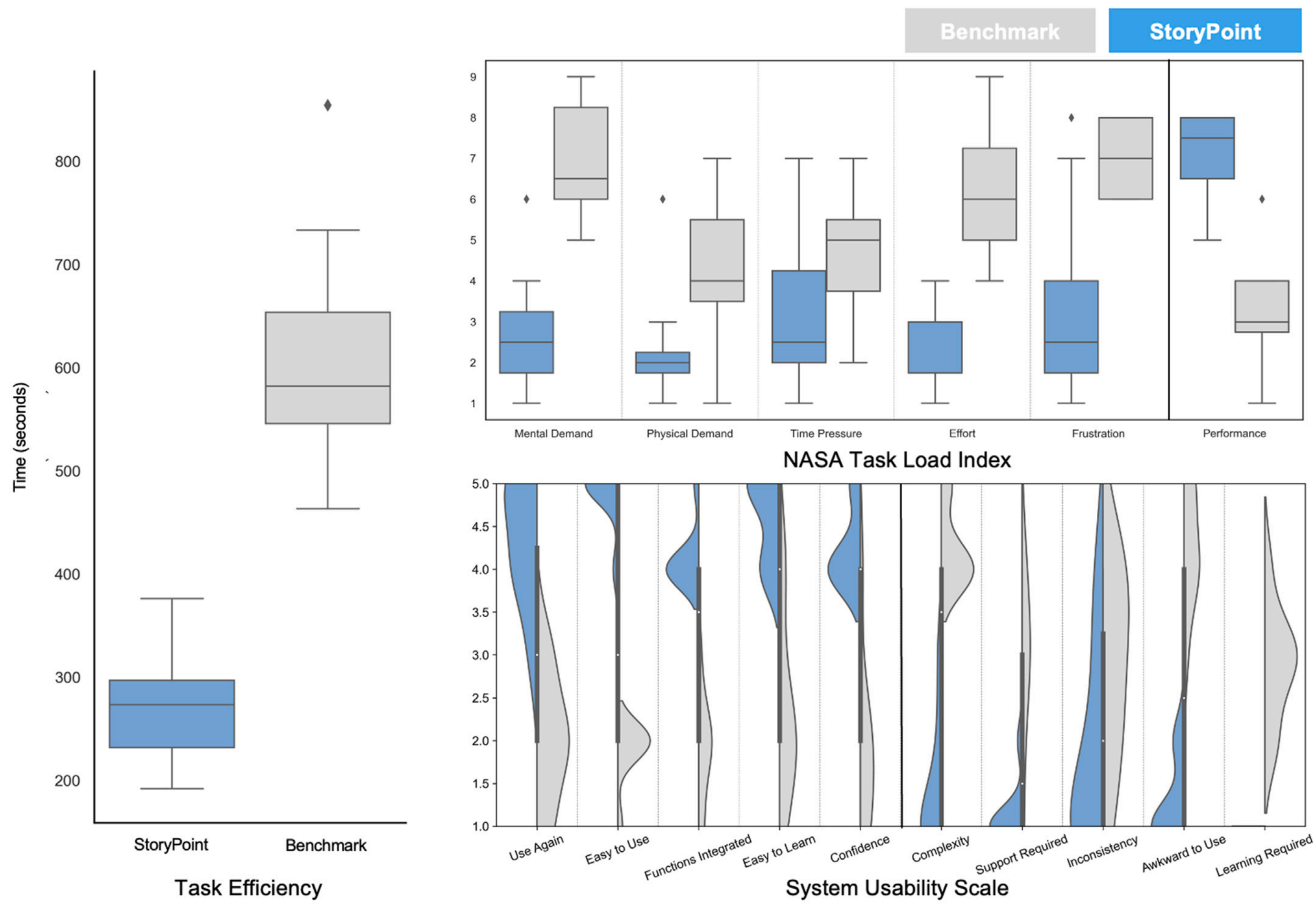

Fig. 6. This figure presents the evaluation results of *StoryPoint* compared to the benchmark tool, showing improved task efficiency with reduced completion time and lower cognitive load. Usability scores indicate that participants found *StoryPoint* easier to use, more integrated, and less frustrating, demonstrating its effectiveness for creating data stories.

that participants found *StoryPoint* (SUS score = 76.875, SD = 5.7863) to be more usable than the benchmark for the given tasks (SUS score = 19.6875, SD = 7.2503). Further statistical details can be found in Fig. 6.

5.2. Evaluation of the quality of the data stories created with StoryPoint

In this evaluation, we aimed to evaluate the quality of the data stories created with *StoryPoint* and compare it with the benchmark. We conducted a power analysis before the study using the stats module from scipy (Virtanen et al., 2020) to estimate an appropriate sample size. Simulations indicated that about 86 participants were needed to achieve an alpha level of 0.05 and a power of 0.8 when using a Mann–Whitney U test for ordinal data. Therefore, we initially recruited 120 participants from Prolific who speak English as their first or second language. Participants were paid $12.5 per hour and provided with a bonus of $0.5 for each correctly completed task (6 tasks in total). To ensure response quality, we embedded three attention checks within the experiment (e.g., specific instructions like “Select option B” and tasks like “What is a data story not?”) and excluded participants who failed more than one check. Additionally, we analyzed task completion times to flag abnormally short or long durations and reviewed responses for inconsistencies or careless patterns. After this filtering process, the final dataset consisted of 104 participants (54 female, 48 male, and 2 non-binary). Their ages ranged from 18 to 73 (M = 37.3431, SD = 11.71). We assessed participants’ proficiency in Python, dashboard utilizing as well as dashboard creating skills using a Likert scale ranging from 1 to 7. On average, the participants demonstrated a low proficiency in Python (M = 2.91), utilizing dashboards (M = 3.42), and creating dashboards (M = 3.26). The median time required by participants for the study was 4 min and 14 s (04:14).

5.2.1. Method

Employing a complete between-subject design, we divided participants into two groups (Group 1: using a random data story created with *StoryPoint*, Group 2: using a data story created with the benchmark) and assigned them tasks. To ensure comparability of these data stories, we drew at random from those created in the first part of our summative evaluation. Participants of each group were tasked with three assignments with increasing complexity: Task 1 involved identifying the most common reason for cart abandonment, task 2 required participants to discern the purpose of the data story, and task 3 necessitated an assessment of customer satisfaction distributions based on age. Each group was exposed to either a random data story from *StoryPoint* or the benchmark in Section 5.1. During the tasks, we meticulously measured the time participants required for each assignment and cross-referenced their results for accuracy. Following the completion of tasks, participants engaged in a subjective evaluation process. They provided feedback on narrative elements, utilizing scales adapted from van der Lee et al. (2021): fluency (1–5), informativeness (1–5), readability (1–5), and overall quality (1–5). Additionally, participants evaluated the quality of the data story using scales derived from Ajani et al. (2022): aesthetics (1–5), clarity (1–5), professionalism (1–5), and trustworthiness (1–5).

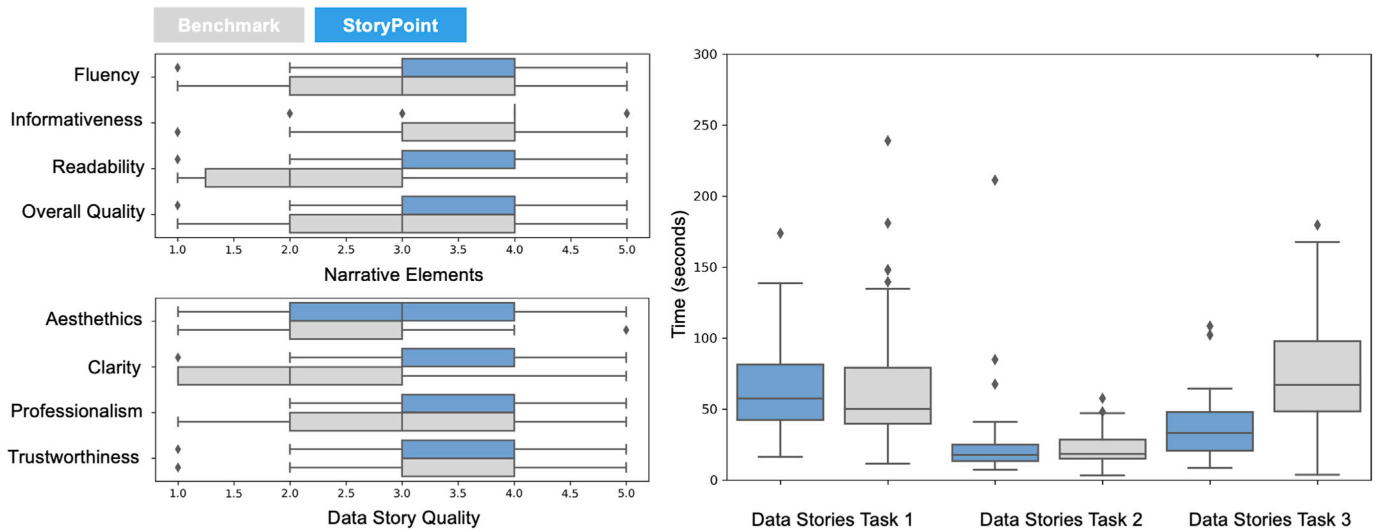

Fig. 7. This figure shows that *StoryPoint* outperforms the benchmark in data story quality, including narrative elements like fluency and readability, and visual aspects such as aesthetics and clarity. Participants further solved the tasks more efficient, highlighting *StoryPoint*'s ability to support creation of high-quality data stories.

5.2.2. Results

We computed p-values using Student's t-test for normally distributed data (time), the two-sample z-test for propositions for correctness, and the Mann–Whitney U test for non-normally distributed data (narrative elements and data story quality) using the stats module from SciPy (Virtanen et al., 2020). The two-sample z-test for propositions revealed a significant effect of the condition on **correctness** ($p = 0.0001$, $z = 15.4684$), indicating that participants were more correct in answering questions with the *StoryPoint* data story (correct responses = 36) than with Baseline data story (correct responses = 25). Concerning the **time** needed, the t-test showed a significant effect of the condition on time for task 3 ($p = 0.0000$, $t = -5.1821$). For the first and second tasks, participants spent a similar amount of time answering questions with the *StoryPoint* data story (M Task 1 = 23.77 s, M Task 2 = 67.03 s) compared to the benchmark data story (M Task 1 = 22.49 s, M Task 2 = 111.32 s). Regarding **narrative elements**, the Mann–Whitney U test indicated that the fluency, informativeness, readability, and overall quality of the narrative were significantly higher with *StoryPoint* compared to the benchmark data story. In terms of overall **data story**, the Mann–Whitney U test reported significantly higher aesthetics, clarity, professionalism, and trustworthiness. For further statistical details on narrative elements and overall data story quality, refer to Table 5 and Fig. 7.

Lastly, we aimed to investigate whether participants' ability to solve tasks correctly influenced their perceptions of narrative elements and data story quality. To achieve this, we conducted additional analyses by filtering participants based on their task performance. **First**, we filtered out 27 participants who solved no tasks correctly. This ensured that remaining participants had a baseline understanding for meaningful evaluation. Mann–Whitney U tests revealed that these participants rated StoryPoint's data stories significantly higher across fluency ($U = 654.0$, $p = 0.0019$), informativeness ($U = 648.0$, $p = 0.0022$), readability ($U = 692.0$, $p = 0.0003$), and overall quality ($U = 617.5$, $p = 0.011$). Additionally, StoryPoint outperformed in aesthetics ($U = 638.0$, $p = 0.0048$), clarity ($U = 702.0$, $p = 0.0001$), professionalism ($U = 642.0$, $p = 0.0035$), and trustworthiness ($U = 617.0$, $p = 0.0095$). **Second**, we analyzed the subset of participants who solved no tasks correctly. This group likely struggled to understand the content of data stories created with either system, making it challenging to identify meaningful differences. Mann–Whitney U tests showed no significant differences in fluency ($U = 134.5$, $p = 0.6652$), informativeness ($U = 169.5$, $p = 0.5017$), readability ($U = 200.5$, $p = 0.0991$), or overall quality ($U = 165.5$, $p = 0.5973$), though clarity was marginally significant favoring StoryPoint ($U = 223.5$, $p = 0.0301$). These findings emphasize the role of comprehension in evaluating the outcomes of data storytelling tools. Participants who solved at least one task correctly consistently favored StoryPoint's data stories, while those without baseline understanding showed no significant preferences. This underscores the need for ensuring comprehension when assessing storytelling tool quality.

5.3. Evaluation of StoryPoint with three real-world use cases

To assess the effectiveness of *StoryPoint* as a data story authoring tool in real-world applications, we recruited three domain experts from diverse enterprise domains, each encompassing a unique enterprise data visualization use case.

5.3.1. Method

In **use case 1**, a sales manager, unaccustomed to data visualization tools such as Tableau or PowerBI, eagerly explored the tool's potential to create a data story containing product sales information. This scenario highlighted the user-friendly nature of *StoryPoint*, making it accessible to non-technical users and underscoring its relevance in diverse professional settings. In **use case 2**, a risk analyst sought to harness the capabilities of our tool to merge historical state-based conflict data with various country-specific key indicators. The objective was to create a data story for assessing the likelihood of future conflict outbreaks. Lastly, **use case 3** featured a marketing manager accustomed to collaborating with data analysts to create and analyze reports. They relied purely on *StoryPoint* to create a data story containing social media key performance indicators. Following the completion of their respective use cases, we inquired about their likelihood to utilize *StoryPoint* for future data story creation. Fig. D.8 shows the data stories created.

5.3.2. Results

For **use case one**, the sales manager, not familiar with typical data visualization tools, uploaded a sales dataset into *StoryPoint*. Despite initial confusion, they quickly adapted after reading the instructions. They entered the prompt *"I'm interested in Sales, Product, Market, Customer, Time"* for the first visualization, expressing satisfaction with the generated bar chart and stating, *"Product [x] has sales of [y]..."* (DE21). The participant found *StoryPoint* to be simple yet impactful, stating, *"Wow, this is really simple and great - I could imagine all my*

Table 5
This table compares *StoryPoint*'s data stories to data stories of the benchmark tool, showing statistically significant improvements in narrative elements (e.g., fluency, informativeness) and data story quality (e.g., aesthetics, clarity).

Item	M StoryPoint	SD StoryPoint	M benchmark	SD benchmark	U-statistics	Significance
Narrative Elements						
Fluency	3.3556	0.9331	2.8793	1.1251	1621.5	**0.0286** $<$ **0.05** ✓
Informativeness	3.8889	0.8318	3.2759	1.0562	1743.0	**0.0020** $<$ **0.05** ✓
Readability	3.3333	1.0660	2.4310	1.1257	1851.0	**0.0001** $<$ **0.05** ✓
Overall Quality	3.4222	0.9412	2.9828	1.1469	1596.5	**0.0441** $<$ **0.05** ✓
Data Story Quality						
Aesthetics	3.0667	1.2136	2.4915	1.1044	1683.0	**0.0165** $<$ **0.05** ✓
Clarity	3.3333	1.0000	2.3898	1.1893	1912.5	**0.0001** $<$ **0.05** ✓
Professionalism	3.7778	0.9266	3.1017	1.0288	1810.0	**0.0009** $<$ **0.05** ✓
Trustworthiness	3.6667	0.9045	3.2373	0.8777	1686.5	**0.0123** $<$ **0.05** ✓

colleagues using the tool" (DE21). Their interaction continued smoothly, with the participant refining prompts, expressing astonishment at *StoryPoint*'s capabilities, stating, *"Even if I took five minutes myself to write such a story purpose, I couldn't have come up with a better one"* (DE21). They confirmed the story purpose and expressed contentment with the resulting data story, noting, *"I really like the result, and it did not take long to create it"* (DE21), expressing an intent to use the tool for future presentations.

For **use case two**, the participant, assessing conflict fatalities for risk evaluation, efficiently used *StoryPoint*'s exploratory data analysis, noting, *"This feature is very helpful; even though I've previously performed explorative data analysis, I haven't experienced such an interactive UI"* (DE22). In the data story creation phase, they entered the prompt *"Plot the data of state-based conflict casualties"* and refined the visualization with commands like *"change month to year in the x-axis name"* and *"also include the data of column state-based conflict casualties sum over the last 24 months in the chart for all countries in the column name, put a second y-axis on the left for this data"*. They expressed satisfaction with the resulting visualizations, stating, *"Sometimes, the easiest narratives are the best"* (DE22). Testing the tool further, they issued complex prompts, and while content with the outcomes, opted to maintain consistency with the initial visualization, stating, *"Okay, this was fast and exactly what I wanted"* (DE22).

In the **third use case**, a marketing manager delved into *StoryPoint* for social media marketing analysis, expressing initial excitement: *"I am interested to see if this tool will have more value for me in comparison to the tools I usually use"* (DE23). They entered the prompt *"Show me the development of impressions over time in a linear graph"* and swiftly customized the resulting chart, stating, *"I'm surprised about how good the tool converted my text to a chart; I wouldn't have thought that it would work so fast"* (DE23). The participant seamlessly progressed to the second visualization, entering prompts such as *"Plot a pie chart that might be interesting in the context of conflict-related fatalities in state-based conflicts"*. They further tested the tool with prompts like *"Calculate the sum of fatalities of state-based conflicts for one year across all countries and depict it as a bar chart"*. Satisfied with the outcomes, they commended *StoryPoint*'s user-friendly layout, stating, *"It looks great and it was easy to use for my use case, I guess I might do that in the future on my own"* (DE23). The participant suggested tool adaptations for specific business areas, stating, *"If the tool could create videos for my marketing campaigns that would solve all my problems"* (DE23).

6. Discussion

Our formative and summative evaluations illustrated that *StoryPoint*, enables domain experts to intuitively create compelling data stories. The results of the user studies show that *StoryPoint*'s data story creation features, such as natural language input alongside pre-calculated charts and narrative elements are effective in providing the right kind of support to domain experts. Participants in our summative evaluation were able to create a data story for their data visualization use case in less than 10 min by following *StoryPoint*'s suggestions and structure.

6.1. Natural language data story creation

During our formative study, we gained valuable insights into the significance of incorporating natural language inputs for chart creation within *StoryPoint*. This feature has proven to be pivotal, as it empowers domain experts to seamlessly craft charts without the necessity of intricate analytical skills. Kavaz et al. (2023) draws the conclusion that the utilization of natural language input empowers users to effortlessly create charts by simply describing their desired outcome, obviating the necessity to specify chart types, columns, and axes explicitly. For instance, domain experts can ask for a visualization of revenue differences between two product lines, instead of asking for a bar chart, with some data field on the y-axis, and grouped by another data field. However, it became apparent that while this freedom of language input was empowering for those without analytical experience, it could also be overwhelming for domain experts already accustomed to creating charts descriptively. The inclusion of customization options within *StoryPoint* raises the question of how to strike a balance between accommodating domain experts' creative needs and providing guidance to ensure optimal data story outcomes. In response to their feedback, we sought to bridge this gap by providing them with illustrative example prompts and concrete example charts, offering guidance on how to generate charts effectively. This approach allowed them to refine and tailor these examples using natural language input, ultimately resulting in a more goal-oriented and intuitive process for domain experts when creating data stories. In this line, we see *StoryPoint* also as a valuable resource for rapid data story prototyping allowing decision-making processes to be streamlined and data-driven insights to be rapidly implemented. Therefore, one intriguing prospect is automating the data story creation process to a greater extent. While *StoryPoint* currently empowers domain experts to create data stories, it would be interesting to explore the possibility of automating this process further, potentially based on user-defined templates or GenAI-driven recommendations. As discussed by Shi et al. (2021) it is important to note that a fully automated approach to data story creation is currently not a feasible option. Therefore the described approach could expedite the creation of standardized data stories while still allowing for manual intervention when nuanced storytelling is required.

6.2. Quality of created data stories

Further, an intriguing aspect of *StoryPoint* that emerges for consideration is the consistency of LLM outputs. As highlighted by one participant who noted that *"when the tool doesn't provide the expected output, one can become frustrated"* (DE15) it becomes evident that the reliability and predictability of LLM-generated content are essential factors influencing effectiveness in the enterprise context. This consistency is particularly significant concerning the trustworthiness of the output, which is a critical requirement for domain experts who rely on accurate results in their decision-making processes. Tam et al. (2022) similarly emphasized the significance of enhancing the factual consistency of

LLMs via summarization, highlighting its crucial importance for end users. Following van der Lee et al. (2021), to enhance the factual consistency of LLMs through summarization, we have incorporated a warning text beneath the text field that narratives generated by ChatGPT may not be fully conclusive. Further, our user studies showed that a compelling data story is not solely dependent on the visualizations it incorporates; the accompanying textual narratives play a crucial role in communicating insights. We learned that users liked the option to rely on automatically generated textual narratives. Therefore, it is imperative to focus on quality assurance measures for both the generated narratives and the resulting data stories. We discovered that narrative characteristics, such as informativeness and readability, and data story quality characteristics, including professionalism and trustworthiness, are crucial for measuring the quality of data stories in the enterprise context (van der Lee et al., 2021; Ajani et al., 2022). In this line, Shi et al. (2021) explored the possibility of training a custom natural language generation (NLG) model to enhance the quality of the generated narratives further. Ensuring the reliability and correctness of the visualizations and narratives produced by *StoryPoint* becomes a central consideration, underlining the tool's effectiveness in meeting the needs of its users.

6.3. Integration of StoryPoint's design features into data visualization tools

We believe that data visualization tools like PowerBI and Tableau can benefit from incorporating the design principles outlined in Table 3. Our user studies showed that more value for domain experts in enterprises could be provided by supporting beyond data visualization to encompass the entire enterprise data storytelling process, assisting users in structuring narratives, and delivering impactful data stories. Next, enhancing existing data visualization tools with natural language input features for chart generation can streamline the process while guiding domain experts toward aligned visualizations. Furthermore, offering predefined data story layouts, as suggested in our research, can encourage users to create uniform and engaging narratives. Incorporating automatically generated explanations, such as overarching data story purposes and chart descriptions, as proposed, aids viewer comprehension. Adding presentation features like an "AskData" (Setlur et al., 2016) function and full-screen mode, which our studies found to be valuable, enhances real-time interactions and immersive experiences. Lastly, integrating audience feedback mechanisms, a key finding from our research, allows viewers to provide input on specific elements of the data story, fostering iterative improvements and effective communication within organizations. By incorporating these principles, PowerBI and Tableau would better empower domain experts and enterprises in their data storytelling endeavors, facilitating more insightful and engaging presentations of data-driven insights.

6.4. Biases in the data storytelling process

Our evaluations revealed several biases for enterprise data storytelling, which we now discuss and propose strategies to mitigate according to literature. Biases in AI-assisted data storytelling can stem from human and algorithmic factors, requiring targeted strategies to ensure fairness and accuracy (Habib and ElTarabishi, 2024). First, **training data bias**, where AI models prioritize common patterns in their datasets, was evident in our system as it occasionally marginalized unique perspectives. We addressed this by grounding narratives more directly in the source data, prompting models for data-aligned outputs, and adding contextual information to the prompts. Validation steps comparing outputs to the original data could further reduce risks of misrepresentation (Fang et al., 2023). Second, **overreliance on GenAI** is a critical concern, particularly in enterprise decision-making contexts, where unverified narratives could lead to poor or even deceptive conclusions (Moller et al., 2018). To counter this risk, *StoryPoint* deliberately adopts a human-in-the-loop design. All generated narratives and charts are presented as drafts, clearly labeled as AI-generated, and accompanied by warnings that outputs may be incomplete or inconclusive. Our evaluations showed that participants consistently verified and refined the outputs, underscoring that human judgment remains central. This ensures that *StoryPoint* operates as a decision-support tool rather than a decision-making system. Additionally, **prompt engineering bias**, stemming from ambiguous or culturally nuanced user inputs, was occasionally apparent in our evaluations. We addressed this by implementing preprocessing mechanisms to standardize prompts, ensuring clarity and neutrality. Finally, **pre-defined templates and generated charts**, though effective for standardization, were sometimes seen as limiting creative exploration.

6.5. Limitations and future work

StoryPoint is not without its limitations and areas for future improvement. Currently, the tool offers limited **customization options** for data stories, such as the inability to rearrange elements or adjust their sizes. While this simplicity may benefit some users, others may desire more flexibility for customizing data stories after creation. Current **query times** (3–5 s for GPT-3.5, up to 10 s for GPT-4) may be a concern for users who expect faster responses. Furthermore, *StoryPoint* lacks advanced **data cleaning** capabilities, leaving data preparation tasks to be handled elsewhere. During the formative evaluation, it became evident that while *StoryPoint* offers greater benefits to domain experts, other user groups such as data analysts may not find it as advantageous. Furthermore, the **export functionality** is limited, and supporting **advanced chart types** beyond the current range of Vega-Lite options, such as tree maps, remains a desirable future enhancement. Finally, the data stories created with *StoryPoint* may exhibit **similarities to dashboards**, influenced by the familiarity of domain experts with dashboard-like interfaces. However, we recognize that this might not embody the optimal structure for organizing data stories in enterprises. Acknowledging this as a current limitation, we emphasize the potential for future research to explore innovative approaches for enhancing narrative structure and elements (Gunklach et al., 2023), guiding users toward these possibilities as needed.

During our evaluation sessions, we received numerous suggestions from users who expressed their intention to incorporate *StoryPoint* into their professional workflows. Building upon this positive feedback, it becomes evident that *StoryPoint* also warrants a longitudinal evaluation. In a proposed field study, we recommend assessing not only the tool itself, but also the broader concept of data stories and authoring tools for business domain experts. This comprehensive evaluation should encompass both qualitative assessments of user perceptions and experiences and quantitative measurements of success. On this basis, we will continue to improve our system to make it viable for enterprise usage. We enthusiastically encourage others to join the open-source development community of *StoryPoint*.

7. Conclusion

Data storytelling as an effective means of communicating data-driven insights is on the rise. However, current data visualization tools offer little assistance in crafting compelling data stories. In this paper, we introduced *StoryPoint*, a data story authoring tool tailored for domain experts in enterprises. *StoryPoint* supports the enterprise data storytelling process, empowering domain experts in combining charts and narrative elements utilizing natural language inputs, to ultimately tell compelling data stories. StoryPoint further demonstrates the potential of integrating domain expertise into GenAI workflows, offering a new direction for data storytelling tools. By embedding contextual metadata and user-defined goals into the data storytelling process, it ensures that charts and narratives are aligned with enterprise purposes. To design *StoryPoint*, we followed a user-centered and literature-supported approach. Our formative evaluation with eight domain experts guided the development process and demonstrated how *StoryPoint* enables rapid prototyping of data stories. A summative evaluation encompassing (i) a small-scale experiment comparing *StoryPoint*'s data creation process with a benchmark using eight domain experts, (ii) a large-scale online experiment with 104 crowdworkers assessing *StoryPoint*'s data story quality, and (iii) three real-world industry data visualization cases, underscores *StoryPoint*'s utility. The future of *StoryPoint* holds promising avenues for further refinement, addressing the nuances of chart generation, enhancing textual narratives, and automating data story creation. We have released *StoryPoint* as an open source project, to enable widespread usage, as well as community-driven improvements. We hope that enterprises and non-profit organizations will implement *StoryPoint within their infrastructure, accompanied by structured training programs to ensure effective use and minimize risks.* Furthermore, our discussion outlines how *StoryPoint*'s design principles could be instantiated into existing commercial data visualization tools like PowerBI and Tableau, enhancing these tools to better support domain experts in creating compelling data stories and presenting their insights more effectively.

CRediT authorship contribution statement

Jonas Gunklach: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elias Mueller:** Writing – review & editing, Writing – original draft, Validation, Software, Conceptualization. **Merlin Knaeble:** Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Alexander Maedche:** Writing – review & editing, Writing – original draft, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Interview questionnaire

1. General Questions About the Participant's Role and Daily Tasks

1. What is your role in the company, and in which area do you work?
2. What are your recurring tasks in your daily work?
3. How familiar are you with the processes in your domain, and where does this knowledge come from?
4. What tools and data sources do you work with regularly?
5. How do you typically use data to support your decision-making processes?

2. Experience and Problems with Data Visualization Tools (e.g., Tableau, PowerBI)

1. What experience do you have with data visualization tools like Tableau or PowerBI? Where did you gain this experience?
2. For what purposes do you typically use these tools, and what are your main objectives?
3. How much time do you usually spend creating and analyzing visualizations?
4. What do you consider the strengths of current data visualization tools?
5. What weaknesses or challenges have you encountered when using these tools?
6. Have you ever had difficulties understanding data stories? What were these issues about, and how did you resolve them?
7. How do you evaluate whether the contents of data stories and the underlying data are trustworthy?
8. What do you do when you need additional or more in-depth information from a dashboard?
9. Have you used tools with AI-powered features like Tableau's "AskData"? What has your experience been like?

3. Requirements for a Data Story Authoring Tool

1. What do you already know about data storytelling?
2. What challenges do you see in communicating data within your organization?
3. How would you imagine an ideal tool for creating data stories?
4. What features would be particularly helpful to better communicate data?
5. Do you think a data story authoring tool could help you present data more effectively? Please explain.
6. What features do you feel are missing from current tools to (1) create data stories easier, (2) present data clearer and easier to understand?
7. Do you think these requirements are generalizable or specific to your business domain?
8. How important would it be for a tool to adapt stories for different audiences automatically?
9. Would features like automated insights, trend highlighting, or contextual summaries be helpful?

4. Think-Aloud Session with Tableau (Including "AskData")

1. For this session, please create a simple data story using Tableau and its "AskData" feature. While doing this, please describe your thoughts out loud.
2. How did you find the process of creating a data story with Tableau?
3. What difficulties did you encounter while using the "AskData" feature?
4. How intuitive did you find the interaction with Tableau during this task?
5. Were there any functionalities missing that would have made creating the data story easier?
6. In your opinion, what could improve the process of creating data stories in tools like Tableau?

Appendix B. Search string

"data storytelling" OR "data-driven storytelling" OR "data stories" OR "data story" OR "narrative visualization" OR "data comics" OR "data journalism" OR (("storytelling" OR "narrative") AND ("data visualization" OR "information visualization" OR "visual analytics" OR "business intelligence"))*

Appendix C. Related work

See Table C.6

Appendix D. Third summative evaluation

See Fig. D.8.

Data availability

We have shared the link to our code in the manuscript footnote.

Table C.6
This table compares related data storytelling tools based on domain knowledge integration, output format, target users, and enterprise integration. It highlights differences in support for domain expertise and interactivity including limited enterprise integration.

Tool	Domain knowledge integration	Output format	Target user	Enterprise integration
Calliope (Shi et al., 2021)	Limited to data-to-chart transformation	Static charts with captions	Non-technical users	Limited integration
Datashot (Wang et al., 2020)	None	Static fact sheets	General users	Partial integration
Text-to-Viz (Cui et al., 2019)	None	Static infographics	Lay users	No integration
ChartAccent (Ren et al., 2017)	Manual annotations only	Annotated static visualizations	Data analysts	No integration
SketchStory (Lee et al., 2013)	User-driven customization	Freeform visuals	Presenters	No integration
Erato (Sun et al., 2022)	User-defined keyframes, collaborative editing	Interactive storylines	Analysts, designers	No integration
Notable (Li et al., 2023)	None	Interactive dashboard narratives	Business analysts	Limited integration
CLUE (Gratzl et al., 2016)	Manual templates for domain knowledge	Static reports	Data analysts	Limited integration
InsideInsights (Mathisen et al., 2019)	Interactive analysis with user-defined narratives	Interactive visuals	General users	No integration

Fig. D.8. This figure shows real-world data stories created with StoryPoint, including analyses of geopolitical conflict trends (use case 1), risk evaluation (use case 2), and social media marketing effectiveness (use case 3).

References

Abbasi, Ahmed, Sarker, Suprateek, University of Virginia, Aalto University, Chiang, Roger, University of Cincinnati, 2016. Big data research in information systems: Toward an inclusive research agenda. JAIS 17 (2), I–XXXII. http://dx.doi.org/10.17705/1jais.00423.

Ajani, Kiran, Lee, Elsie, Xiong, Cindy, Knaflic, Cole Nussbaumer, Kemper, William, Franconeri, Steven, 2022. Declutter and focus: Empirically evaluating design guidelines for effective data communication. IEEE Trans. Vis. Comput. Graph. 28 (10), 3351–3364. http://dx.doi.org/10.1109/TVCG.2021.3068337.

Amini, Fereshteh, Brehmer, Matthew, Bolduan, Gordon, Elmer, Christina, Wiederkehr, Benjamin, 2018. Evaluating data-driven stories and storytelling tools. In: Data-Driven Storytelling. AK Peters/CRC Press, pp. 249–286.

Anuradha, J., et al., 2015. A brief introduction on Big Data 5Vs characteristics and Hadoop technology. Procedia Comput. Sci. 48, 319–324.

Baumer, Eric P.S., Jasim, Mahmood, Sarvghad, Ali, Mahyar, Narges, 2022. Of course it's political! a critical inquiry into underemphasized dimensions in civic text visualization. In: Computer Graphics Forum, vol. 41, (no. 3), Wiley Online Library, pp. 1–14.

Boldosova, Valeriia, Luoto, Severi, 2019. Storytelling, business analytics and big data interpretation: Literature review and theoretical propositions. MRR 43 (2), 204–222. http://dx.doi.org/10.1108/MRR-03-2019-0106.

Boldosova, Valeriia, Luoto, Severi, 2020. Storytelling, business analytics and big data interpretation: Literature review and theoretical propositions. Manag. Res. Rev. 43 (2), 204–222.

Brehmer, Matthew, Kosara, Robert, 2021. From jam session to recital: Synchronous communication and collaboration around data in organizations. IEEE Trans. Vis. Comput. Graphics 28 (1), 1139–1149.

Brooke, John, 1996. Sus: a "quick and dirty'usability. Usability Eval. Ind. 189 (3), 189–194.

Bryant, Antony, Charmaz, Kathy, 2007. The Sage Handbook of Grounded Theory. Sage.

Chao, Gong, 2009. Human-computer interaction: process and principles of human-computer interface design. In: 2009 International Conference on Computer and Automation Engineering. IEEE, pp. 230–233.

Chen, Qing, Cao, Shixiong, Wang, Jiazhe, Cao, Nan, 2023. How does automation shape the process of narrative visualization: A survey of tools. IEEE Trans. Vis. Comput. Graphics.

Chevalier, Fanny, Tory, Melanie, Lee, Bongshin, van Wijk, Jarke, Santucci, Giuseppe, Dörk, Marian, Hullman, Jessica, 2018. From analysis to communication: Supporting the lifecycle of a story. In: Data-Driven Storytelling. AK Peters/CRC Press, pp. 151–183.

Crisan, Anamaria, Fiore-Gartland, Brittany, Tory, Melanie, 2020. Passing the data baton: A retrospective analysis on data science work and workers. IEEE Trans. Vis. Comput. Graphics 27 (2), 1860–1870.

Cui, Weiwei, Zhang, Xiaoyu, Wang, Yun, Huang, He, Chen, Bei, Fang, Lei, Zhang, Haidong, Lou, Jian-Guan, Zhang, Dongmei, 2019. Text-to-viz: Automatic generation of infographics from proportion-related natural language statements. IEEE Trans. Vis. Comput. Graphics 26 (1), 906–916.

Elias, Micheline, Aufaure, Marie-Aude, Bezerianos, Anastasia, 2013. Storytelling in visual analytics tools for business intelligence. In: Human-Computer Interaction, Vol. 8119. INTERACT 2013, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 280–297. http://dx.doi.org/10.1007/978-3-642-40477-1_18.

Fang, Xiaoxuan, Ng, Davy Tsz Kit, Leung, Jac Ka Lok, Chu, Samuel Kai Wah, 2023. A systematic review of artificial intelligence technologies used for story writing. Educ. Inf. Technol. 28 (11), 14361–14397.

Gratzl, Samuel, Lex, Alexander, Gehlenborg, Nils, Cosgrove, Nicola, Streit, Marc, 2016. From visual exploration to storytelling and back again. In: Computer Graphics Forum, vol. 35, (no. 3), Wiley Online Library, pp. 491–500.

Gunklach, Jonas, Jacob, Katharina, Michalczyk, Sven, 2023. Beyond dashboards? Designing data stories for effective use in business intelligence and analytics. In: ECIS 2023 Research Papers. 327.

Gunklach, Jonas, Nadj, Mario, Knaeble, Merlin, Bragaglia, Isabela, Maedche, Alexander, 2024. Designing for effective human-guided machine learning feasibility analysis. In: ICIS 2024 Proceeding. 1.

Gunklach, Jonas, Nadj, Mario, Michalczyk, Sven, Jacob, Katharina, Gröger, Christoph, Mädche, Alexander, 2025. Beyond the unicorn? Job roles in data science. Bus. Inf. Syst. Eng. 1–25.

Habib, Mary Y., ElTarabishi, Maha, 2024. The role of Artificial Intelligence in shaping the future of Media production and the application of Algorithm bias theory in Storytelling. J. Media Interdiscip. Stud. 3 (8), 1–28.
Hart, Sandra G., 2006. NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, (no. 9), Sage publications Sage CA, Los Angeles, CA, pp. 904–908.
Haug, Saskia, Benke, Ivo, Maedche, Alexander, 2023. Aligning crowdworker perspectives and feedback outcomes in crowd-feedback system design. Proc. ACM Hum.-Comput. Interact. 7 (CSCW1), 1–28. http://dx.doi.org/10.1145/3579456.
He, Yi, Cao, Shixiong, Shi, Yang, Chen, Qing, Xu, Ke, Cao, Nan, 2024. Leveraging large models for crafting narrative visualization: a survey. arXiv preprint arXiv:2401.14010.
Holzinger, Andreas, Kargl, Michaela, Kipperer, Bettina, Regitnig, Peter, Plass, Markus, Müller, Heimo, 2022. Personas for artificial intelligence (AI) an open source toolbox. IEEE Access 10, 23732–23747.
Hu, Kevin, Orghian, Diana, Hidalgo, César, 2018. DIVE: A mixed-initiative system supporting integrated data exploration workflows. In: Proceedings of the Workshop on Human-in-the-Loop Data Analytics. ACM, Houston TX USA, pp. 1–7. http://dx.doi.org/10.1145/3209900.3209910.
Huang, Xin, Li, Xiang, Yao, Jianlin, 2023. Investigating the role of visual storytelling in enhancing audience engagement in digital journalism. MEDAAD 2023, 10–16.
Hullman, Jessica, Diakopoulos, Nick, 2011. Visualization rhetoric: Framing effects in narrative visualization. IEEE Trans. Vis. Comput. Graphics 17 (12), 2231–2240.
Hutchinson, Maeve, Jianu, Radu, Slingsby, Aidan, Madhyastha, Pranava, 2024. LLM-assisted visual analytics: Opportunities and challenges. arXiv preprint arXiv:2409.02691.
Jung, Soon-Gyo, Salminen, Joni, Aldous, Kholoud Khalil, Jansen, Bernard J., 2025. Personacraft: Leveraging language models for data-driven persona development. Int. J. Hum.-Comput. Stud. 103445.
Kandel, Sean, Paepcke, Andreas, Hellerstein, Joseph M., Heer, Jeffrey, 2012. Enterprise data analysis and visualization: An interview study. IEEE Trans. Vis. Comput. Graphics 18 (12), 2917–2926.
Kavaz, Ecem, Puig, Anna, Rodríguez, Inmaculada, 2023. Chatbot-based natural language interfaces for data visualisation: A scoping review. Appl. Sci. 13 (12), 7025. http://dx.doi.org/10.3390/app13127025.
Kim, Hyeok, Srinivasan, Arjun, Brehmer, Matthew, 2024. Bringing data into the conversation: Adapting content from business intelligence dashboards for threaded collaboration platforms. In: 2024 IEEE Visualization and Visual Analytics. VIS, IEEE, pp. 81–85.
Lee, Bongshin, Kazi, Rubaiat Habib, Smith, Greg, 2013. SketchStory: Telling more engaging stories with data through freeform sketching. IEEE Trans. Vis. Comput. Graphics 19 (12), 2416–2425.
Lee, Bongshin, Riche, Nathalie Henry, Isenberg, Petra, Carpendale, Sheelagh, 2015. More than telling a story: Transforming data into visually shared stories. IEEE Comput. Grap. Appl. 35 (5), 84–90. http://dx.doi.org/10.1109/MCG.2015.99.
Lennerholt, Christian, Van Laere, Joeri, Söderström, Eva, 2021. User-related challenges of self-service business intelligence. Inf. Syst. Manage. 38 (4), 309–323. http://dx.doi.org/10.1080/10580530.2020.1814458.
Li, Haotian, Wang, Yun, Qu, Huamin, 2024. Where are we so far? understanding data storytelling tools from the perspective of human-ai collaboration. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–19.
Li, Haotian, Wang, Yun, Qu, Huamin, 2025. Reflection on data storytelling tools in the generative AI era from the human-AI collaboration perspective. arXiv preprint arXiv:2503.02631.
Li, Haotian, Ying, Lu, Zhang, Haidong, Wu, Yingcai, Qu, Huamin, Wang, Yun, 2023. Notable: on-the-fly assistant for data storytelling in computational notebooks. In: CHI '23: CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, pp. 1–16. http://dx.doi.org/10.1145/3544548.3580965.
Liem, J., Perin, C., Wood, J., 2020. Structure and empathy in visual data storytelling: evaluating their influence on attitude. Comput. Graph. Forum 39 (3), 277–289. http://dx.doi.org/10.1111/cgf.13980.
Lin, Yanna, Yang, Leni, Li, Haotian, Qu, Huamin, Moritz, Dominik, 2025. Interlink: Linking text with code and output in computational notebooks. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–15.
Logeshwaran, J., Srivastava, Durgesh, Kumar, K. Sree, Rex, M. Jenolin, Al-Rasheed, Amal, Getahun, Masresha, Soufiene, Ben Othman, 2024. Improving crop production using an agro-deep learning framework in precision agriculture. BMC Bioinformatics 25 (1), 341.
Martinez-Maldonado, Roberto, Echeverria, Vanessa, Fernandez Nieto, Gloria, Buckingham Shum, Simon, 2020. From data to insights: a layered storytelling approach for multimodal learning analytics. In: CHI '20: CHI Conference on Human Factors in Computing Systems. ACM, Honolulu HI USA, pp. 1–15. http://dx.doi.org/10.1145/3313831.3376148.
Masson, Damien, Malacria, Sylvain, Casiez, Géry, Vogel, Daniel, 2023. Charagraph: interactive generation of charts for realtime annotation of data-rich paragraphs. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, pp. 1–18. http://dx.doi.org/10.1145/3544548.3581091.
Mathisen, Andreas, Horak, Tom, Klokmose, Clemens Nylandsted, Grønbæk, Kaj, Elmqvist, Niklas, 2019. Insideinsights: Integrating data-driven reporting in collaborative visual analytics. In: Computer Graphics Forum, vol. 38, (no. 3), Wiley Online Library, pp. 649–661.
Microsoft, 2024. Overview of copilot for power BI - Power BI. https://learn.microsoft.com/en-us/power-bi/create-reports/copilot-introduction.
Microsoft Corporation, 2025. Power BI: Copilot integration. https://learn.microsoft.com/en-us/power-bi/create-reports/copilot-integration. (Accessed 22 September 2025).
Moller, Judith, Trilling, Damian, Helberger, Natali, van Es, Bram, 2018. Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. Inf. Commun. Soc. 21 (7), 959–977.
Norman, Donald A., 1983. Design principles for human-computer interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1–10.
OECD, 2019. An OECD learning framework 2030. Futur. Educ. Labor 23–35.
OECD, 2020. Curriculum Overload: A Way Forward. OECD, http://dx.doi.org/10.1787/3081ceca-en.
Ramm, Saskia, Kopf, Eva-Maria, Dinter, Barbara, Hönigsberg, Sarah, 2021. What makes a good story - the use and acceptance of storytelling in business intelligence. In: Hawaii International Conference on System Sciences. http://dx.doi.org/10.24251/HICSS.2021.692.
Ren, Donghao, Brehmer, Matthew, Lee, Bongshin, Höllerer, Tobias, Choe, Eun Kyoung, 2017. Chartaccent: Annotation for data-driven storytelling. In: 2017 IEEE Pacific Visualization Symposium. PacificVis, Ieee, pp. 230–239.
Ren, Pengkun, Wang, Yi, Zhao, Fan, 2023. Re-understanding of data storytelling tools from a narrative perspective. Vis. Intell. 1 (1), 11. http://dx.doi.org/10.1007/s44267-023-00011-0.
Ritchie, Jane, Lewis, Jane, Nicholls, Carol McNaughton, Ormston, Rachel, 2013. Qualitative Research Practice: A Guide for Social Science Students and Researchers. sage.
Ruoff, Marcel, Myers, Brad A., Maedche, Alexander, 2023. ONYX: assisting users in teaching natural language interfaces through multi-modal interactive task learning. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, pp. 1–16. http://dx.doi.org/10.1145/3544548.3580964.
Sarikaya, Alper, Correll, Michael, Bartram, Lyn, Tory, Melanie, Fisher, Danyel, 2019. What do we talk about when we talk about dashboards? IEEE Trans. Vis. Comput. Graph. 25 (1), 682–692. http://dx.doi.org/10.1109/TVCG.2018.2864903.
Satyanarayan, Arvind, Moritz, Dominik, Wongsuphasawat, Kanit, Heer, Jeffrey, 2016. Vega-lite: A grammar of interactive graphics. IEEE Trans. Vis. Comput. Graphics 23 (1), 341–350.
Segel, E., Heer, J., 2010. Narrative visualization: telling stories with data. IEEE Trans. Vis. Comput. Graph. 16 (6), 1139–1148. http://dx.doi.org/10.1109/TVCG.2010.179.
Setlur, Vidya, Battersby, Sarah E., Tory, Melanie, Gossweiler, Rich, Chang, Angel X., 2016. Eviza: A natural language interface for visual analysis. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology. pp. 365–377.
Shao, Hongbo, Martinez-Maldonado, Roberto, Echeverria, Vanessa, Yan, Lixiang, Gasevic, Dragan, 2024. Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension? In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–21.
Shi, Danqing, Xu, Xinyue, Sun, Fuling, Shi, Yang, Cao, Nan, 2021. Calliope: Automatic visual data story generation from a spreadsheet. IEEE Trans. Vis. Comput. Graph. 27 (2), 453–463. http://dx.doi.org/10.1109/TVCG.2020.3030403.
Showkat, Dilruba, Baumer, Eric P.S., 2021. Where do stories come from? examining the exploration process in investigative data journalism. Proc. ACM Hum.-Comput. Interact. 5 (CSCW2), 1–31.
Sun, Mengdi, Cai, Ligan, Cui, Weiwei, Wu, Yanqiu, Shi, Yang, Cao, Nan, 2022. Erato: Cooperative data story editing via fact interpolation. IEEE Trans. Vis. Comput. Graphics 29 (1), 983–993.
Tableau, 2024. Tableau AI: Innovate for the future with generative AI. https://www.tableau.com/learn/webinars/tableau-gpt-innovate-future-generative-ai-emea.
Tableau, 2025. Tableau pulse. https://www.tableau.com/de-de/products/tableau-pulse. (Accessed 22 September 2025).
Tam, Derek, Mascarenhas, Anisha, Zhang, Shiyue, Kwan, Sarah, Bansal, Mohit, Raffel, Colin, 2022. Evaluating the factual consistency of large language models through summarization. arXiv preprint arXiv:2211.08412.
Tory, Melanie, Bartram, Lyn, Fiore-Gartland, Brittany, Crisan, Anamaria, 2022. Finding their data voice: Practices and challenges of dashboard users. IEEE Comput. Grap. Appl. http://dx.doi.org/10.1109/MCG.2021.3136545, 1–1.
Trieu, Van-Hau, Burton-Jones, Andrew, The University of Queensland, Green, Peter, Queensland University of Technology, Cockcroft, Sophie, University of Southern Queensland, Springfield Campus, 2022. Applying and extending the theory of effective use in a business intelligence context. MISQ 46 (1), 645–678. http://dx.doi.org/10.25300/MISQ/2022/14880.
van der Lee, Chris, Gatt, Albert, van Miltenburg, Emiel, Krahmer, Emiel, 2021. Human evaluation of automatically generated text: Current trends and best practice guidelines. Comput. Speech Lang. 67, 101151.
Virtanen, Pauli, Gommers, Ralf, Oliphant, Travis E., Haberland, Matt, Reddy, Tyler, Cournapeau, David, Burovski, Evgeni, Peterson, Pearu, Weckesser, Warren, Bright, Jonathan, et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. Nature Methods 17 (3), 261–272.

Wang, Yun, Sun, Zhida, Zhang, Haidong, Cui, Weiwei, Xu, Ke, Ma, Xiaojuan, Zhang, Dongmei, 2020. DataShot: Automatic generation of fact sheets from tabular data. IEEE Trans. Vis. Comput. Graph. 26 (1), 895–905. http://dx.doi.org/10.1109/TVCG.2019.2934398.

Watson, Hugh J., 2017. Data visualization, data interpreters, and storytelling. Bus. Intell. J. 22 (1), 6.

Webster, Jane, Watson, Richard T., 2002. Analyzing the past to prepare for the future: Writing a literature review. MIS Q. xiii–xxiii.

Wongsuphasawat, Kanit, Moritz, Dominik, Anand, Anushka, Mackinlay, Jock, Howe, Bill, Heer, Jeffrey, 2016. Voyager: exploratory analysis via faceted browsing of visualization recommendations. IEEE Trans. Vis. Comput. Graphics 22 (1), 10.

Wu, Aoyu, Wang, Yun, Zhou, Mengyu, He, Xinyi, Zhang, Haidong, Qu, Huamin, Zhang, Dongmei, 2022. MultiVision: Designing analytical dashboards with deep learning based recommendation. IEEE Trans. Vis. Comput. Graph. 28 (1), 162–172. http://dx.doi.org/10.1109/TVCG.2021.3114826.

Ye, Yilin, Hao, Jianing, Hou, Yihan, Wang, Zhan, Xiao, Shishi, Luo, Yuyu, Zeng, Wei, 2024. Generative ai for visualization: State of the art and future directions. Vis. Inform..

Zhao, Jian, Xu, Shenyu, Chandrasegaran, Senthil, Bryan, Chris, Du, Fan, Mishra, Aditi, Qian, Xin, Li, Yiran, Ma, Kwan-Liu, 2023. ChartStory: Automated partitioning, layout, and captioning of charts into comic-style narratives. IEEE Trans. Vis. Comput. Graph. 29 (2), 1384–1399. http://dx.doi.org/10.1109/TVCG.2021.3114211.