Received: 18 December 2024 | Revised: 21 October 2025 | Accepted: 28 October 2025

DOI: 10.1002/asi.70036

RESEARCH ARTICLE

JASIST WILEY

Using generative AI to co-design data-driven stories

Angelica Lo Duca

Institute of Informatics and Telematics, National Research Council, Pisa, Italy

Correspondence
Angelica Lo Duca, Institute of Informatics and Telematics, National Research Council, Pisa, Italy.
Email: angelica.loduca@iit.cnr.it

Abstract

Data storytelling combines communication and visualization techniques with storytelling principles to convey insights meaningfully and engagingly. This paper introduces the AI-DIKW framework, an extension of the Story Data–Information–Knowledge–Wisdom (S-DIKW) hierarchy, designed to support the co-creation of data-driven stories by combining human expertise with Generative AI. The AI-DIKW framework extends the S-DIKW model through the application of the journalistic 5-Ws, which guide Generative AI to provide consistent and contextually relevant outputs. This study describes how Generative AI can contribute to the iterative construction of a story's key components while ensuring alignment with ethical principles. Applying Generative AI at the AI-Data stage involves extracting insights from the data. Applying Generative AI at the AI-Information stage means enriching the insight with relevant context, while at the AI-Knowledge stage, Generative AI helps the data storyteller to enrich the story with meaningful next steps, anchored to a selected ethical framework. Finally, at the AI-Wisdom stage, Generative AI helps the data storyteller to tailor the story to specific audiences. At each stage of the AI-DIKW framework, Generative AI acts as a co-designer that helps the data storyteller to frame and edit the story. The article also describes a case study that applies the concepts described practically.

1 | INTRODUCTION

Information science organizes, processes, and communicates data, information, and knowledge. As information environments become increasingly complex, new technologies, including Generative Artificial Intelligence (AI), offer transformative potential in how insights are extracted, structured, and presented (Gozalo-Brizuela & Garrido-Merchan, 2023). Generative AI has already been applied to different domains, including education (Zapata-Rivera et al., 2024) and information communication in healthcare (Clay et al., 2024). One area where this transformation is still underexplored is data storytelling, which integrates data visualization, narrative techniques, and audience engagement strategies to convey insights meaningfully. Data storytelling is a powerful way to communicate insights extracted from data by transforming them into narratives (Dykes, 2019). More formally, data storytelling builds compelling stories extracted from data, allowing analysts and data scientists to present and share their insights interestingly and interactively through a story. Good data storytelling requires a mix of art and science. Art comes in finding the right story, while science understands how to use data to support that story (Cairo, 2012).

This paper situates Generative AI-assisted data storytelling within an information science framework, demonstrating how AI can enhance information structuring, representation, and dissemination in data-rich environments. The role of narrative in information science has been increasingly recognized as a crucial mechanism for organizing and communicating complex data (Matei & Hunter, 2021). Storytelling facilitates knowledge transfer, comprehension, and decision-making, offering an alternative to traditional, static data presentations (Lee et al., 2015; Segel & Heer, 2010). At the same time, data visualization has evolved as a core component of information science, with researchers exploring how visual encoding can enhance human cognition (Franconeri et al., 2021). By integrating storytelling and data visualization, data storytelling provides a framework for transforming raw data into meaningful insights tailored to specific audiences.

The application of AI to data storytelling aligns with long-standing discussions in information science about automation, augmentation, and knowledge representation (Pasquinelli & Joler, 2021). Li et al. have already analyzed the possible ways of using Generative AI in data storytelling (Li et al., 2023). They defined four potential uses of Generative AI: creator, optimizer, reviewer, and assistant. Using Generative AI as a creator means that Generative AI generates all the content. As an optimizer, Generative AI fine-tunes the content generated by humans. When acting as a reviewer, Generative AI evaluates the human-generated content by identifying potential issues and proposing possible solutions. Finally, when acting as an assistant, Generative AI helps humans to complete their tasks.

While prior research has examined AI's role in natural language processing, data extraction, and visualization (He et al., 2024) with a focus on isolated tasks rather than comprehensive methodologies (Lo Duca, 2023; Lo Duca, 2024a), there is a need for systematic methods that integrate AI into the entire data storytelling pipeline. This paper addresses that gap by introducing the AI-DIKW framework, which enriches the well-established Data-Information-Knowledge-Wisdom (DIKW) model with Generative AI to facilitate data-driven storytelling.

The DIKW framework is central to information science, serving as a conceptual bridge between data processing and human decision-making (Van Meter, 2020). DIKW has been widely used in fields such as knowledge management, digital libraries, and decision support systems to structure information through progressive layers of meaning. However, traditional DIKW implementations often rely on manual curation and interpretation, which can limit scalability. The S-DIKW framework is a specific application of the DIKW framework to data storytelling (Lo Duca & McDowell, 2025; McDowell, 2021). The AI-DIKW framework extends this paradigm by integrating Generative AI at each stage, using AI to extract insights from data, enrich information with context, provide actionable next steps, and tailor the story to diverse audiences. In doing so, it positions Generative AI as a cognitive assistant rather than a replacement for human storytelling, aligning with broader discussions in information science about the augmentation of human expertise through computational tools (Dykes, 2024).

This paper extends the S-DIKW framework proposed by Lo Duca and McDowell (2025) and advances the field of information science and technology in several ways. First, it provides a structured methodology for integrating Generative AI into data storytelling, grounding this process in the S-DIKW framework. Building on this foundation, the paper introduces the AI-DIKW framework, which draws on the journalistic 5-Ws (Singer, 2008) as examples of tasks that AI systems can use to support reasoning at different stages of data-story construction, while acknowledging that they represent only one possible way AI can contribute to data storytelling. The journalistic 5-Ws can be understood as guiding principles to ensure that AI outputs remain consistent, contextually relevant, and adaptable to specific audiences. Beyond the methodological dimension, the paper also connects the framework to tool design, highlighting how AI-DIKW can inform the development and refinement of data storytelling platforms, whether through slides, infographics, or interactive formats.

To demonstrate the effectiveness of this approach, the paper discusses a case study that applies the AI-DIKW framework to climate change data, tailoring stories for different audience types. The case study uses OpenAI ChatGPT (OpenAI, 2024a) as a Generative AI tool.

By aligning data storytelling with core information science principles, such as knowledge organization, user engagement, and information retrieval, this research extends existing discussions on the role of AI in shaping modern information environments. Rather than positioning AI as a replacement for human creativity or judgment, this approach highlights AI's potential to augment human capabilities, making storytelling more efficient and adaptable (Dykes, 2024).

The remainder of the paper is organized as follows: section 2 describes the related literature about data storytelling and the application of AI to it. Section 3 proposes the AI-DIKW framework, and section 4 applies it to the case study. Finally, section 5 proposes conclusions and future work.

2 | RELATED WORK

2.1 | Defining data storytelling

Data storytelling can be broadly defined as the practice of combining data, visuals, and narrative to communicate insights in a way that is both comprehensible and persuasive (Dykes, 2019). Unlike traditional data visualization, which focuses primarily on graphical representation, data storytelling integrates narrative elements that guide interpretation and highlight meaning. Lee et al. (2015) emphasize the importance of annotations, labels, and textual components for effective storytelling, while Knaflic (2015) frames it as an enhanced form of data visualization that facilitates understanding. Franconeri et al. (2021) and Pandey et al. (2014) underline its persuasive dimension, pointing to the rhetorical power of visuals.

Scholars have also explored the narrative foundations of data storytelling. Segel and Heer (2010) identify seven genres of narrative visualization, while Matei and Hunter (2021) argue that an effective data-driven story must not only report facts but also challenge assumptions through causal explanation. Other contributions apply established narrative frameworks from literature and cinema, such as Freytag's Pyramid (Yang et al., 2022), the Hero's Journey (Wei et al., 2025), or the Three-Act Structure (Lo Duca, 2025). In this view, data storytelling emerges as a hybrid form of communication that blends empirical evidence with narrative techniques to enhance comprehension, persuasion, and memorability.

Building on these conceptualizations, data storytelling has been applied in multiple domains, each with distinct objectives. In business communication, data storytelling is used to deliver the insights extracted from data to help decision-making (Dykes, 2019). In education, data storytelling is used either to help teachers assess the quality of learners (educational data storytelling) (Milesi & Martinez-Maldonado, 2024; Sarica & Yildirim, 2024) or to help learners understand data (Matuk et al., 2022). Finally, data journalism uses data storytelling to equip journalistic articles with data and narratives (Ojo & Heravi, 2018; Yu & Stasko, 2024).

2.2 | Artificial intelligence for data storytelling and visualization

The application of AI to data storytelling and visualization is still new in the literature, although there is an increasing interest in the scientific community. He et al. conducted an extensive survey of 79 papers on how AI is applied to crafting narratives at different stages (He et al., 2024), including data, narration, visualization, and presentation. For a complete review, refer to their work. This paper highlights only some representative papers. Li et al. experimented to understand if and how much story creators could use AI in their stories (Li et al., 2023; Li et al., 2024). The experiment demonstrated that there is still a certain reluctance due to a lack of knowledge about AI tools. Lo Duca built a preliminary AI-assisted data storytelling tool to help story creators transform their data into stories (Lo Duca, 2023; Lo Duca, 2024b). Other works focused on using AI for specific tasks, such as title (Liu et al., 2023), caption (Liew & Mueller, 2022), or annotation (Chen & Liu, 2023) generation, while creating visualizations. While storytelling may be possible to automate at a surface level, AI-assisted storytelling will require human judgments to reach and engage audiences effectively (Brooks et al., 2023).

AI is also used in data visualization to help the data analysis process. Wang et al. proposed three levels of integration between AI and data visualization to extract knowledge (Wang et al., 2023). At Level 0, AI and data visualization work separately; at Level 1, two strategies can be defined: AI4VIS (Wang et al., 2021) and VIS4AI (Yuan et al., 2021). At Level 3, AI and data visualization work together to generate knowledge. Compared to these works on data visualization, this paper does not address the problem of using data visualization for data analysis. Instead, this paper focuses on how to build data-driven stories starting from data and how to communicate them effectively.

Compared to the existing literature, this paper defines a methodology incorporating Generative AI in the human process of crafting data-driven stories. Following the existing literature, this paper uses Generative AI as an assistant, which augments human knowledge and does not replace it. In addition, this paper assumes that human judgment is always included in AI-assisted story building.

3 | THE AI-DIKW FRAMEWORK

The DIKW framework is a conceptual mindset that provides macro steps to transform data into wisdom, following other intermediate steps, which include information and knowledge. The DIKW framework is independent of data storytelling and can be used as a general framework for conceptual data representation. This paper uses the elements of the DIKW framework as progressive steps to transform data into data stories. This idea is not new in data storytelling; Berengueres and Sandell proposed this approach (Berengueres & Sandell, 2019), and Lo Duca and McDowell improved it by detailing the main elements required to build a data-driven story (Lo Duca & McDowell, 2025; McDowell, 2021). More specifically, Lo Duca and McDowell (2025) proposed the S-DIKW framework, which combines the DIKW framework and the classical narrative arc, and details the visual elements required at each step to transform a data visualization into a data story. Table 1 summarizes the S-DIKW framework.

TABLE 1 The S-DIKW framework proposed by Lo Duca and McDowell.

S-DIKW element	Description	Mapping to the narrative arc
S-Data	Describe the insight extracted from the data	Second Act
S-Information	Add context to the story to make it understandable	First Act
S-Knowledge	Suggest the next steps to follow when the story ends	Third Act
S-Wisdom	Adapt the story to the audience	Story Editing

An alternative interpretation might consider insights as information extracted from data, insights plus context as knowledge, and calls for action as the locus of wisdom (Lo Duca, 2024a). Audience adaptation could then be seen as a separate editorial phase. This perspective is coherent with other readings of DIKW, and it offers a valuable lens for further refinement of the framework. In this paper, the sequence is maintained as proposed in the S-DIKW framework, since it supports an iterative co-design process with Generative AI and reflects the practical workflow adopted in the case study. However, acknowledging alternative mappings enriches the theoretical robustness of the model.

Conceptually, the S-DIKW framework is similar to the Data Storytelling Arc proposed by Dykes (2019) and more detailed by Lo Duca (2025), which identifies three main elements in a story: the beginning of the story, the main point, and the end. Presenting insights from data corresponds to the main point of a data story. Adding context to the extracted information corresponds to defining the story's beginning, with background, and raising the audience's interest. Adding a call to action corresponds to the last part of the story and invites the audience to take action. The described mapping suggests the natural flow when building a data story. First, the story's main point should be identified, then the background set, and finally, the next steps should be defined. This flow is precisely what is determined by the S-DIKW framework. When the story is ready, it can be adapted to the specific audience type (Longnecker, 2023).

In the classical narrative, before writing a story, the message to be conveyed should be defined. In the case of data storytelling, however, the narrative must emerge from the data itself: the process starts from the insight that the data reveal and then communicates what is discovered (Lo Duca, 2025). The S-DIKW framework supports this bottom-up approach, ensuring that the story is anchored in evidence rather than imposed from pre-defined assumptions. This perspective distinguishes data storytelling from "storytelling with data," which often implies a top-down approach and carries the risk of cherry-picking information to fit a predetermined message.

This paper improves the S-DIKW framework by adding the assistance of Generative AI. Generative AI is used as a story co-designer that helps data storytellers extract insights from the data and organize them into a story. Generative AI does not replace the data storyteller. Instead, it acts as an augmented brain that helps them delve into the data. The resulting framework, called AI-DIKW, comprises the following steps:

- AI-Data uses Generative AI to extract insight from the data.
- AI-Information uses the advanced features of Generative AI, such as AI agents and deep research, to retrieve information about context.
- AI-Knowledge uses Generative AI to propose next steps based on some (ethical) criteria proposed by humans.
- AI-Wisdom uses Generative AI to organize the elements extracted from the previous steps as a story and adapt the tone and the language of the story to different audiences.

It is important to emphasize that the AI-DIKW framework is not a fully automated process. Rather, human intervention is required after each step to ensure thorough analysis, critical filtering, and careful selection of the outputs produced by Generative AI. The co-design process between humans and Generative AI unfolds as a conversational interaction, in which each stage of the framework (from AI-Data to AI-Wisdom) is executed through one or more instructions, or prompts. This dialogic approach benefits both participants: the data storyteller progressively advances from data toward wisdom, refining their understanding through the outputs generated during the interaction and wisely interpreted, while Generative AI incrementally enriches its contextual understanding of the data as the conversation evolves. This process is inherently iterative, as the data storyteller may need to move back and forth across the different stages of the AI-DIKW framework to gather additional evidence or to validate emerging intuitions.

The AI-DIKW framework adopts one possible operationalization of the DIKW hierarchy for data storytelling. It builds on the S-DIKW model (Lo Duca & McDowell, 2025), where each level is reinterpreted as a narrative function. Other mappings between DIKW and storytelling activities are possible and can be equally valid, depending on the theoretical perspective adopted.

To guide Generative AI in producing meaningful outputs within the AI-DIKW framework, this study refers to the journalistic 5-Ws (Who, What, When, Where, Why) (Singer, 2008). Originally used as a heuristic for structuring news narratives, the 5-Ws are employed here as one possible scaffold to inspire and orient the AI's reasoning process at different stages.

3.1 | AI-Data

In the literature, different definitions of insight have been proposed, including a unit of discovery, a collection of knowledge, a hypothesis or evidence, data facts, or knowledge links (Battle & Ottley, 2023). This paper considers insight as something relevant that emerges from the data, such as an anomaly, a relationship between the data, a trend, and so on (Berengueres & Sandell, 2019). There are different techniques to extract an insight from the data, spanning from curiosity-driven analysis (Klein, 2013) to more deterministic approaches (Ojo & Heravi, 2018). Table 2 provides a synthesis of six analytical techniques commonly employed for insight extraction from data, each mapped to one of the 5-Ws of journalism to highlight their investigative orientation. For each technique, a brief definition is provided along with practical strategies that can be operationalized in data analysis workflows. The techniques span temporal, spatial, categorical, and relational perspectives, as well as anomaly detection and entity-focused exploration. The table serves both as a conceptual framework and a practical reference for selecting the most appropriate analytical method depending on the research question and the structure of the available data.

Data storytellers can employ Generative AI to extract meaningful insights, and propose appropriate graphical representations. The process entails providing the input dataset to the Generative AI system, specifying the technique to be applied (as outlined in Table 2), and instructing Generative AI to implement all relevant strategies for that technique in order to identify and extract all potential insights contained within the data. Data storytellers review the produced output and, optionally, ask Generative AI to apply another technique to extract another type of insight.

The proposed approach presupposes that the Generative AI system receives the complete dataset as input (Bryant & Mukherjee, 2023). However, if the dataset contains sensitive information, it should first undergo a preliminary anonymization process prior to submission. In addition, in cases where the dataset exceeds the

TABLE 2 The six analytics techniques to extract an insight from the data.

Technique	Mapping to 5-Ws	Definition	Strategies for implementation
Temporal Analysis	When	Investigation of changes in a phenomenon over time	Peak detection Extreme values comparison Highlight seasonality Threshold analysis
Spatial Analysis	Where	Evaluation of a phenomenon across different locations	Side-by-side geographic comparison Geospatial visualization Hierarchical geographical scaling Cross-level metric comparison
Multi-Category Comparison	What	Comparison of categories within the same temporal and spatial frame	Dominant category identification Homogeneous group comparison Positional anomaly assessment Paired temporal comparison
Outlier Detection	What/Why	Identification of values or patterns deviating from expected distributions	Point anomalies Collective anomalies Contextual anomalies
Correlation and Relationship Analysis	Why	Examination of associations between two or more variables	Correlation measurement Causality assessment Temporal lag analysis
Entity Analysis	Who	Identification and characterization of key actors in the dataset	Frequency analysis Attribute-based comparison Network mapping

TABLE 3 The six types of context.

Context type	Mapping to 5-Ws	Definition
Historical Background	When	Provide temporal background by linking the data to relevant past events, trends, or developments
Spatial and Geographical Context	Where	Specify the geographical setting of the data and situates the phenomenon within its spatial context
Domain and Background	What	Provide domain-specific background, definitions, and the significance of the data
Comparisons and Benchmarking	What	Define reference points or comparative measures used to evaluate the data's relative position or performance
Causes or Explanations	Why	Explain possible causes, contributing factors, and related events influencing the data
Stakeholders and Actors	Who	Identify the entities or groups involved in or affected by the data

maximum input size permitted by the Generative AI tool, a representative subset of the data should be extracted and provided instead, ensuring that it accurately reflects the characteristics of the entire population.

Once insights have been extracted from the data, the subsequent step is their graphical representation. Generative AI tools specifically trained for code generation, such as GitHub Copilot (GitHub, 2024), or more general-purpose systems, such as OpenAI ChatGPT, can be employed to suggest or produce the code required to create the charts, or to directly generate the visualizations themselves. Prompts provided to the Generative AI must describe the desired output in natural language. In more complex scenarios, prompts should also specify the operations necessary to produce that output. The system then returns either the finalized chart or the source code for constructing it in the chosen programming language. Prior work by Lo Duca (2024a) provides detailed guidance on employing Generative AI tools for the generation of visualization code.

3.2 | AI-Information

Context refers to the surrounding elements that make the audience understand the data story. Adding context to a data story improves the efficiency and effectiveness of insights comprehension (Shao et al., 2024). Conceptually, context can provide different types of information. Table 3 categorizes eight types of contextual information relevant to data-driven storytelling, aligning each with the corresponding "W" from the journalistic framework. For each type, it defines its purpose and outlines strategies for systematic retrieval and integration. These contexts range from domain, temporal, and spatial framing to comparative benchmarks, causal explanations, and stakeholder mapping. Together, they provide a structured guide for enriching raw insights with meaningful information thereby operationalizing the AI-Information phase of the AI-DIKW framework.

Data storytellers can leverage the advanced capabilities of Generative AI, such as deep research functionalities or AI agents (Huang et al., 2025), to retrieve relevant contextual information for the data under analysis. The process begins by providing the Generative AI system with a detailed description of the dataset, the insights obtained from the AI-Data phase, the specific type of context to be explored, the strategies to be applied (as outlined in Table 3), and any additional relevant information. In response, the Generative AI produces a set of contextual outputs, which the data storyteller then reviews, analyzes, and filters. The interaction may proceed iteratively, with the data storyteller requesting further elaboration or clarification on one or more of the generated outputs. This process continues until the resulting context is deemed adequate. It is important to note that, when instructing Generative AI to extract contextual information, the data storyteller should explicitly request the inclusion of links to original sources, thereby facilitating the verification of the retrieved material.

Traditionally context is provided in charts in two ways: textual and visual (Horn, 1998). Textual context includes the title, commentaries, and annotations. The title is the chart title. A commentary is the text that precedes the visualization chart. It consists of the background that helps the audience set the scene and understand the insight. An annotation is a short text explaining a chart's detail, for example, an abnormal point or a trendline. Visual context includes images and symbols. An image is a picture enforcing the commentary or the annotation. Symbols include arrows, circles, lines, and so on, combined with annotations. They help the audience focus on specific points or areas of a chart.

Generative AI can assist data storytellers in improving textual and visual contexts in a data visualization chart, using tools such as OpenAI ChatGPT (OpenAI, 2024a) and DALL-E 3 (OpenAI, 2024b). More specifically, once the data storyteller has identified the type or types of context to be incorporated into the narrative (as described in Table 3), they may instruct the Generative AI to produce a textual or visual context to be integrated into the story. This can be achieved through a single direct prompt or via a sequence of iterative, refinement prompts. During the AI-Wisdom phase described later in the paper, the generated context can then be further tailored to align with the overarching objectives of the narrative.

3.3 | AI-Knowledge

Applied to data storytelling, the knowledge stage empowers individuals to make more informed and effective decisions based on data. At this level, the story, now enriched with contextual information derived from the previous phases of the framework, guides the definition of actionable next steps. In practice, this means that data, once contextualized, is internalized and translated into concrete actions to be implemented. Previous work defined some possible next steps to implement, including learning more, asking for support, freely interacting, providing different options, and proposing a plan (Lo Duca, 2025; Lo Duca & McDowell, 2025). This paper details the strategy to implement next steps practically through the 5-Ws of journalism (Table 4).

TABLE 4 5Ws framework for calls to action.

5-Ws	Role in call to action	Guiding questions
Who	Identify the audience or stakeholders expected to act	Who is the target of this call to action? Who holds responsibility or influence?
What	Define the specific action to be performed	What exactly should the audience do?
When	Establish the time frame or urgency of the action	When should the action be carried out?
Where	Specify the channel, platform, or location where the action should occur	Where should the action take place?
Why	Provide the justification and motivation for taking the action	Why is this action necessary or valuable?

Using the 5-Ws, data storytellers ensure that the proposed action is clear (What), directed (Who), timely (When), situated (Where), and justified (Why). This increases audience comprehension and enhances engagement and the likelihood of action.

At the knowledge stage, data storytellers can employ Generative AI to structure a call to action by addressing the journalistic 5-Ws in a systematic and coherent manner. Beyond answering these fundamental questions, Generative AI can also assist in anchoring a call to action within a specific ethical framework, thereby ensuring that proposed actions are not only practical but also morally grounded. In practice, Generative AI can support the definition of the most appropriate way to design and implement a call to action.

Generative AI functions as a natural interlocutor in this process, enabling data storytellers to refine their reasoning, explore alternatives, and develop actionable strategies through an iterative conversational exchange. This dialogic capacity allows storytellers to deliberate on complex decisions, assess potential consequences, and progressively arrive at more nuanced strategies. At this stage, the ability of Generative AI to engage in interactive dialogue is deliberately exploited as a means of enhancing human judgment and decision-making.

Crucially, good judgment—defined as the capacity to assess situations accurately and to make decisions aligned with values and goals—requires that calls to action be grounded in ethical reflection. Anchoring a story's message to an ethical framework provides a principled basis for evaluating the consequences of actions, safeguarding against misinterpretation, and preventing unintended harm. Such frameworks serve as normative guidelines for determining the best course of action according to principles of fairness, justice, and responsibility.

Several ethical traditions can be applied in this context (Berengueres & Sandell, 2019):

- Utilitarianism evaluates actions by their ability to maximize utility or happiness for the greatest number of people.
- Deontology, which emphasizes adherence to moral duties and rules, holds that some actions are inherently right or wrong regardless of their outcomes.
- Virtue ethics, which prioritizes the cultivation of moral character traits (e.g., courage, honesty, compassion) over strict rule-following or outcome-maximization.

- Care ethics, which highlight empathy, relationships, and responsibility for meeting the needs of others as the foundation of moral action.

Regardless of which ethical framework is selected, any call to action derived from data storytelling should be explicitly anchored within one of them. Doing so ensures that decisions are not made solely on the basis of expediency or immediate effectiveness, but instead are aligned with broader moral principles and their impact on individuals, communities, and society at large. In this way, data storytelling moves beyond the mere presentation of information toward the creation of positive change rooted in ethical considerations. Specifically, Generative AI can be instructed to tailor a call to action so that it aligns with a selected ethical framework, explicitly incorporating the principles of utilitarianism, deontology, virtue ethics, or care ethics as guiding criteria. At the knowledge level of the AI-DIKW framework, storytellers thus leverage the augmented reasoning capabilities of Generative AI to design calls to action that are both actionable and ethically robust, ensuring that decisions derived from data are responsible, meaningful, and justifiable.

At the end of the AI-knowledge stage, the data storyteller has prepared all the material to build the data story, organized into three acts: the First Act, with the context defined at the AI-Information stage, the Second Act, with the insight extracted from the AI-Data stage, and the Third Act, with the next steps defined at the AI-Knowledge stage.

3.4 | AI-Wisdom

The wisdom stage of the AI-DIKW framework focuses on refining the data story so that it is effectively adapted to a specific audience. In this paper, AI-Wisdom refers to the phase in which the story is adapted to specific audiences in terms of tone, language, values and communicative intent. This interpretation follows the original definition of S-Wisdom (Lo Duca & McDowell, 2025), which concerns choosing how, when and with whom a story should be shared. At the same time, wisdom in DIKW also involves responsibility, and awareness of the consequences of communication, dimensions that this model acknowledges and can further accommodate.

Previous studies have broadly distinguished three primary audience categories: the general public, professionals, and executives or decision-makers (Lo Duca, 2025; Lo Duca & McDowell, 2025). Building on these general categories, it is possible to conceptualize audiences more precisely through the lens of the journalistic 5-Ws. An audience corresponds to who the message is directed at; it is embedded within a specific cultural background (what), situated in a particular time (when) and place (where), and motivated by particular values or needs (why).

Adapting a story to an audience therefore requires careful consideration of both tone and language. Tone refers to the emotional register or attitude conveyed in the narrative (e.g., formal vs. informal, urgent vs. neutral), while language encompasses the lexical and syntactic choices that make the story accessible and persuasive for the intended audience. Every data-driven story is created with a specific objective, typically to inform, persuade, or entertain (Dykes, 2024). Aligning this communicative goal with audience type ensures relevance and effectiveness. For instance, the general public may require simplified language and contextual explanations, professionals may expect technical detail and evidence-based reasoning, while executives may prioritize actionable insights and strategic implications.

The data storyteller can leverage the general knowledge of a Generative AI system, or augment it with contextual sources via retrieval-augmented generation (RAG) (Arslan et al., 2024), to customize the tone and language of the story elements derived from the previous steps (context, insight, and call to action) in alignment with audience characteristics. Generative AI can generate textual and visual elements tailored both to the communicative objective and the specific audience profile. In doing so, Generative AI facilitates the adaptation of data stories so that they not only convey insights but also resonate with the values, expectations, and needs of their recipients.

In summary, the AI-DIKW framework operates as a co-design process in which data storytellers and Generative AI collaborate at every stage. The integration of the journalistic 5-Ws offers a possible mechanism to support Generative AI in engaging with key questions that arise throughout the process, helping to encourage systematic exploration and meaningful outcomes.

3.5 | Implications for tool design

The main implication of the AI-DIKW framework for tool design is its modularity. AI-DIKW provides a flexible reasoning structure that can be adopted selectively, depending on the scope and maturity of a tool. Each layer of the framework may function independently or in combination with others, which allows designers to identify the cognitive stages already supported and to recognize where gaps persist. From an implementation point of view, a data story generated through AI-DIKW can take different media forms, according to the needs of the storyteller and the target audience. It may appear as slides for a presentation, an infographic, or info-images (Coelho & Mueller, 2020). The framework does not prescribe a single output format; it offers a backbone that can be embedded in different tool ecosystems.

Using modularity in this way shifts the use of AI-DIKW toward a diagnostic and design-oriented lens. Visualization tools such as REMEX (Li et al., 2025) already support the structuring of meta-relations between data objects but tend to remain within the information or knowledge layer. Integrating AI-DIKW would make it possible to extend these tools toward contextualization and wisdom-level reasoning, where implications and ethical considerations can be articulated. Other tools, such as Melody (Renda et al., 2023), which enables co-design of narratives through Linked Open Data, support contextualization and adaptation to audiences but do not include structured mechanisms for insight generation or next-step reasoning. AI-DIKW makes these limitations explicit and can orient the evolution of such platforms toward layered narrative reasoning.

Other systems, such as tools for annotated visualization design (Lee et al., 2015) and for automated caption or title generation (Liew & Mueller, 2022; Liu et al., 2023), align only partially with the framework. They address the passage from data to information but do not include higher-order interpretation, value-driven framing or guidance for decisions. In these cases, AI-DIKW can indicate where additional cognitive scaffolding may be introduced to complete the reasoning process.

4 | CASE STUDY

This section illustrates how to transform the data related to the Global Temperature Anomalies released by NOAA under the Creative Commons 1.0 Universal Public Domain Dedication (CC0-1.0) license (NOAA, 2024) into a story tailored to two different audiences: decision-makers and the general audience. The story objective for decision-makers will be to help them make decisions, while the purpose of the general audience is to inform them. The raw dataset contains the monthly temperature anomalies from January 1850 to December 2023. Temperature anomalies are with respect to the 20th-century average. This case study is part of the studies on raising awareness of various audiences on climate change. For example, Hine et al. conducted a study on motivating a dismissive, disengaged, and alarmed audience (Hine et al., 2016). They underlined how all the types of audiences analyzed are more involved if the message transmitted contains specific adaptation advice. Figure 1 shows a visual representation of the data.

4.1 | AI-Data

The first step was insight extraction. Since the dataset contains temporal data, ChatGPT was asked to use the temporal technique to extract an insight (derived from Table 2). The following prompt was used for ChatGPT along with the dataset uploaded as a CSV file:

Consider the attached dataset containing Global Temperature Anomalies from 1850 to 2024 released by NOAA. Apply Temporal Analysis (When) to extract insights from the data. Specifically: Investigate how the phenomenon changes over time (identify long-term trends). Perform peak detection to highlight significant maxima or minima. Compare extreme values across different time periods. Highlight seasonality or recurring temporal patterns, if present. Conduct a threshold analysis to determine when the values exceed or fall below critical levels. Summarize the results clearly and concisely in natural language, and present both descriptive statistics and possible visualizations that illustrate the findings.

ChatGPT provided a very detailed answer, including different visual representations of the dataset (Figure 2) and details about descriptive statistics, long-term trends, peaks and valleys, threshold analysis, and seasonality.

As a summary insight, ChatGPT highlighted a pronounced long-term warming trend. The coldest anomalies were observed in the early 20th century, while the warmest occurred in the 21st century, particularly during the 2020s. The recent clustering of years exceeding a $+1^\circ$C anomaly indicates the emergence of a distinct new climate regime. Seasonal variability appears minimal; instead, the defining feature of the dataset is the persistent upward shift in global temperature anomalies. Accordingly, the central insight for the story is the existence of a strong and sustained long-term increase in temperature anomalies.

4.2 | AI-Information

After extracting the insight, the next step of the AI-DIKW framework was asking ChatGPT to add context to the story. Given the complexity of the topic described by the dataset, ChatGPT was instructed to concentrate specifically on domain and background aspects (derived from Table 3). To ensure continuity, the interaction was conducted within the same conversational thread as the AI-Data phase. At this stage, the following prompt was

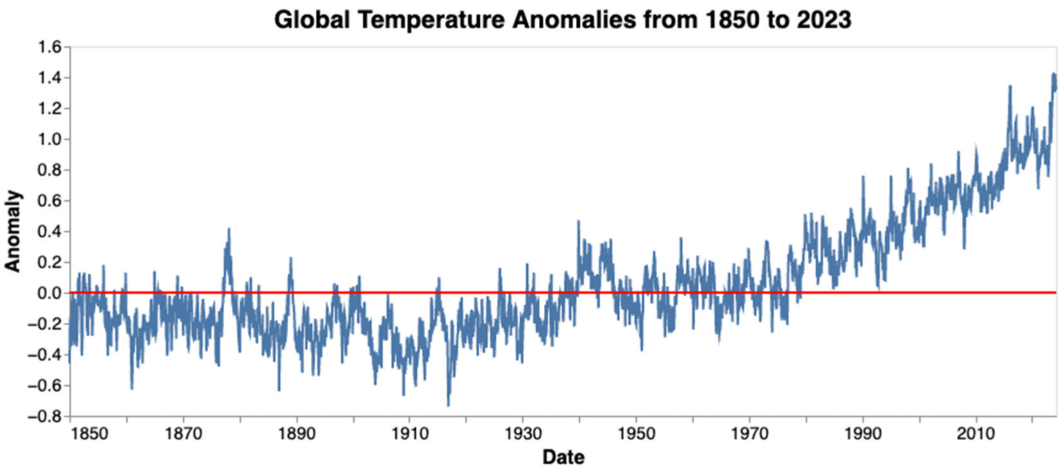

FIGURE 1 Global temperature anomalies from 1850 to 2023. OpenAI ChatGPT (version 5 Fast) was used as a Generative AI tool during the experiments. The detailed output produced by ChatGPT at each stage is available in Data S1. All the prompts were written by the author on the basis of the 5 Ws defined at each step. The outputs generated by ChatGPT were reviewed by the author after each step.

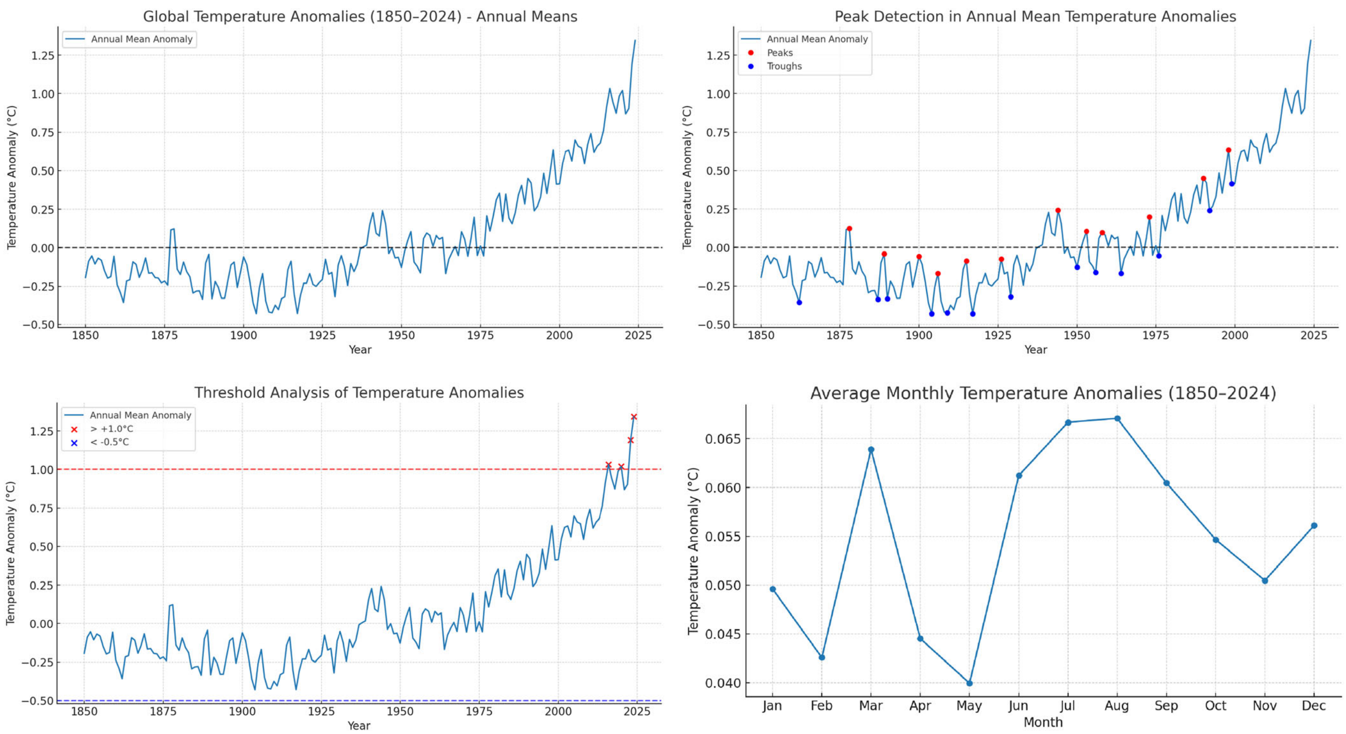

FIGURE 2 The graphs generated by ChatGPT during the AI-Data stage.

submitted to ChatGPT, with the Deep Research functionality activated via the dedicated interface option:

Enrich the analysis by providing domain-specific background and context. Specifically, please: Explain the domain to which the dataset belongs, including key concepts and definitions that are necessary to understand the data. Describe the historical and scientific significance of the data, explaining why it matters and what it represents in its field. Identify relevant background knowledge, such as related events, studies, or standards that help situate the dataset in context. Highlight how this background information can support a better interpretation of the insights previously extracted at the AI-Data level. Include references to credible sources (with links where available) to enable verification and further exploration. Summarize the findings in clear academic language that is accessible to both experts and non-experts.

As output, ChatGPT generated a report addressing multiple dimensions: a domain overview with key concepts, the historical and scientific significance of the data, a background discussion of major developments and standards, and an interpretation informed by domain knowledge. At the domain level, it clarified definitions such as temperature anomalies, baseline periods, and the distinction between climate and weather. It then outlined long-term warming trends, emphasizing the role of anthropogenic climate change alongside natural events such as major volcanic eruptions (e.g., Krakatoa in 1883, Pinatubo in 1991) and strong El Niño episodes (e.g., 1877–78, 1997–98, 2015–16). At the background level, it traced the evolution of climate monitoring from 19th-century scientific efforts to modern datasets and highlighted the establishment of standards and policy targets, including the 1.5°C pre-industrial benchmark. Finally, at the interpretation level, it contextualized warming magnitudes, baseline anomalies, and acceleration of trends, clarified the influence of natural variability, confirmed findings across multiple datasets, and linked the results to policy-relevant thresholds. For this analysis, ChatGPT drew upon diverse sources, including Wikipedia, NASA, Berkeley Earth (an independent U.S.-based nonprofit organization specializing in environmental science and data analysis), and Climate Lab Book (a climate science blog).

4.3 | AI-Knowledge

At this stage, the process advanced to the definition of next steps, framed within an explicit ethical perspective. The action *asking for support* was selected as the next step, grounded in the principles of *care ethics*. To implement this, the following prompt was provided to ChatGPT within the same conversational thread as the AI-Data and AI-Information stages:

Based on the insights extracted from the dataset (AI-Data level) and the contextual information provided (AI-Information level), design the next step in the form of a call to action. The chosen next step is "ask for support," and it must be explicitly grounded in the principles of care ethics. Please: Formulate a call to action that emphasizes empathy, relationships, and responsibility toward others. Explain why support is needed and who should be involved (stakeholders, communities, or institutions). Clarify how this support could positively impact the community or individuals affected by the data insights. Ensure that the language reflects compassion, inclusivity, and relational values, in line with care ethics. Suggest possible channels or strategies to communicate this call to action effectively. Present the response in clear language suitable for data storytelling, and explicitly reference care ethics principles as the ethical grounding.

ChatGPT organized the output at four levels. First, it focused on why support is needed, framing climate change through the lens of care ethics as a relational issue that disproportionately affects vulnerable populations and demands solidarity across generations and geographies. Second, it identified who should be involved, ranging from local communities and grassroots organizations to governments, international institutions, businesses, scientists, and everyday citizens. Third, it detailed how support can help, emphasizing adaptation, mitigation, and relational care as interconnected strategies. Finally, it suggested how to communicate this call, highlighting storytelling, community engagement, collaboration with trusted institutions, and digital campaigns as effective channels. Overall, the message situates support as an expression of mutual responsibility, grounded in compassion and shared care for both people and the planet.

4.4 | AI-Wisdom

At this stage, the core story elements—context, insight, and next steps—are complete and require adaptation to a specific audience. For illustrative purposes, this study develops a data story structured around four components: a title, a textual description of the context, a graph visualizing the extracted insight, and a textual description of the proposed next steps. This case study demonstrates how ChatGPT was employed to generate the textual elements of the story, adapting them to the intended audience. However, the same approach can be followed to generate other types of elements, such as images or short videos. The only component not produced with ChatGPT was the graph itself (Figure 1), although the tool is equally capable of generating such visualizations, as illustrated in Figure 2. To demonstrate how tone and language shift depending on audience characteristics, this paper examines two audience types: executives and the general public. Each audience is defined according to the 5-Ws of journalism, which provide a structured framework for audience characterization.

4.4.1 | Executives

In the example, the audience of executives is composed of senior executives in the renewable energy sector (Who), embedded in a sustainability-oriented cultural background (What), situated within the temporal

TABLE 5 The final story for the audience of executives.

Narrative arc	Story
Title	Data to Action: How Record Temperatures Call for Bold Renewable Investment
First Act (Context)	Since 1850, global temperatures have risen by over 1.3°C, with the last decade the hottest on record. This warming, driven by greenhouse gas emissions, directly impacts economies, societies, and ecosystems. International frameworks such as the Paris Agreement set limits of 1.5–2°C, shaping policy and markets. For the renewable energy sector, this data underscores both the urgency and the opportunity: accelerating clean transitions aligns climate responsibility with profitable growth
Section Act	Figure 1 shows the global temperature anomalies trendline
(Insight)	Global temperatures hit record highs, up 1.3–1.5°C since 1850. The data underline urgent demand growth for renewables—an opportunity for profitable, responsible sector leadership
Third Act (Next Steps)	Possible next steps: 1. Accelerate Investment in Renewables—Scale wind, solar, and storage projects to meet rising demand and achieve 2030 neutrality targets. 2. Strengthen Cross-Sector Partnerships—Collaborate with governments, financiers, and communities to expand clean energy infrastructure. 3. Innovate for Efficiency—Invest in R&D for grid integration, smart systems, and next-gen technologies that improve profitability and resilience. 4. Embed Care-Centered Leadership—Prioritize social responsibility by ensuring fair access, just transitions, and community benefits. 5. Lead Global Advocacy—Champion ambitious climate policies and market standards, positioning the sector as a driver of both growth and sustainability

horizon of the 2030 carbon neutrality goals (When), operating in European and global energy markets (Where), and motivated by the need to balance profitability with environmental responsibility and strategic leadership (Why).

ChatGPT was asked to generate the title, the textual context and the textual next steps, through a conversation, starting with the following prompt (in the same thread of the previous conversation), defining the audience and asking to generate the context.

Consider an audience of senior executives in the renewable energy sector (Who), embedded in a sustainability-oriented culture (What), aligned with the 2030 carbon neutrality goals (When), operating in European and global energy markets (Where), and motivated by profitability, responsibility, and leadership (Why). Select only the background information required by this type of audience to understand the topic and describe it with 500 characters.

The following prompts asked to generate the next steps, a description of the insight and five possible titles. The complete conversation is available in Data S1, Supporting Information. After revising the produced output, the final story described in Table 5 was generated. The story can be implemented as a single infographic, infomage, or multiple slides in a presentation, using the preferred tool and Figure 3 shows a possible visualization implementing it.

4.4.2 | General audience

This audience is composed of citizens concerned with environmental issues (Who), embedded in everyday cultural narratives on climate change (What), situated in the present context of extreme weather events and sustainability debates (When), living in urban European settings influenced by local green policies (Where), and motivated by the desire to protect family well-being, reduce costs, and secure a better future (Why). ChatGPT was guided using the specific characteristics of this audience to respond to the same set of prompts employed for the executive group. The full conversational transcript is provided in Data S1, while Table 6 presents the resulting final story for this type of audience and Figure 4 shows a possible visualization implementing it.

5 | DISCUSSION

The AI-DIKW framework was applied to present insights on global temperature anomalies, tailoring the narrative to two distinct audiences: executives and the general audience. Despite using the same data, the narrative's framing, tone, and style were customized for each audience. For the general public, the text was jargon-free and straightforward, with titles and textual descriptions designed for accessibility. The story

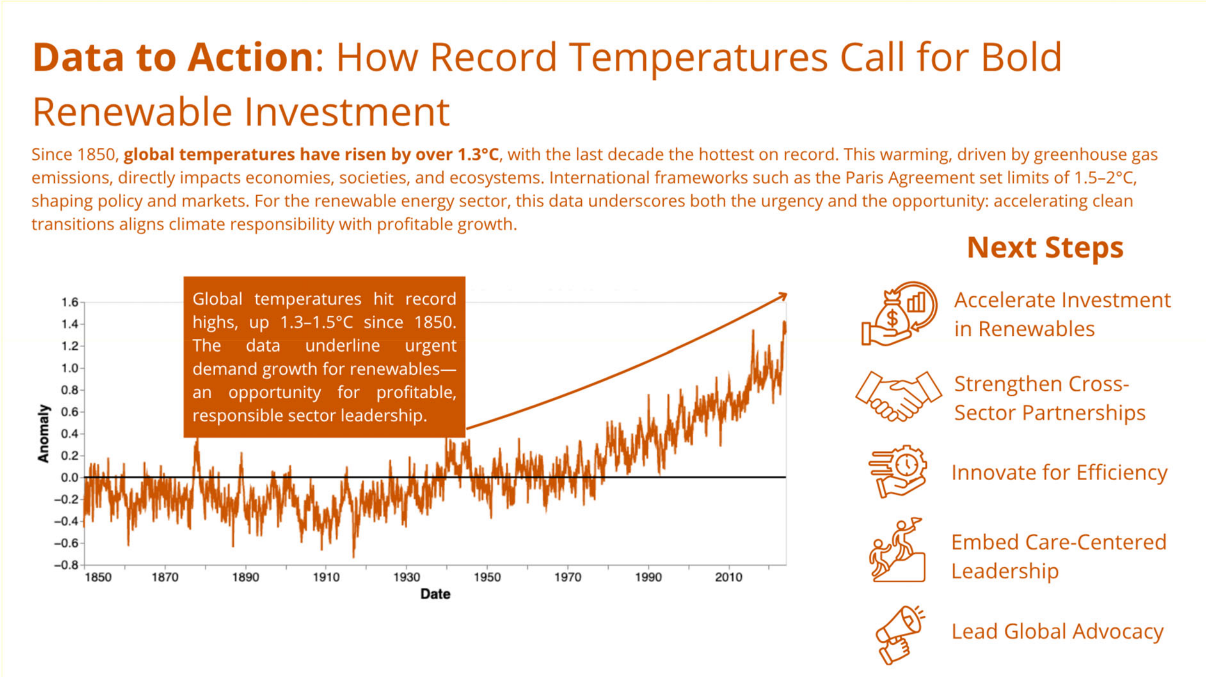

FIGURE 3 The final data story for executives.

TABLE 6 The final story for the general audience.

Narrative arc	Story
Title	Caring for Tomorrow: Responding to Climate Data with Support and Solidarity
First Act (Context)	Global temperatures have risen over 1.3°C since the 1800s, making recent years the hottest on record. This warming drives extreme weather, rising costs, and risks to health and family well-being. Scientists link the trend to greenhouse gases from energy, transport, and industry. Europe's green policies aim to cut emissions and build resilience. Understanding these changes helps citizens see how climate action today safeguards communities, lowers costs, and protects future generations
Section Act	Figure 1 shows the global temperature anomalies trendline
(Insight)	The data shows our planet has warmed over 1.3°C since the 1800s, making recent years the hottest ever. This drives extreme weather, higher costs, and risks for families and communities
Third Act (Next Steps)	Possible next steps: 1. Save Energy at Home—Insulate, switch to efficient appliances, and reduce heating/cooling use. Cuts bills and emissions. 2. Choose Sustainable Transport—Walk, cycle, use public transit, or switch to EVs. Lowers costs and urban air pollution. 3. Adopt a Greener Diet—Eat more plant-based foods and waste less. Supports health and reduces food costs. 4. Support Local Green Policies—Back city initiatives on clean energy, recycling, and green spaces. Strengthens community resilience. 5. Invest in Renewable Choices—Opt for green electricity plans or solar panels. Long-term savings and climate benefits

targeted decision-makers, employing more technical and action-oriented language to convey urgency and strategic importance. Both cases included tailored calls to action. For the general audience, suggestions were practical and straightforward, such as energy-saving tips. For decision-makers, recommendations were strategic and systemic, focusing on renewable energy and policy initiatives.

These case studies demonstrate the flexibility of Generative AI in tailoring data-driven stories to different audiences, effectively aligning the same dataset with varied objectives. The results highlight the potential of the

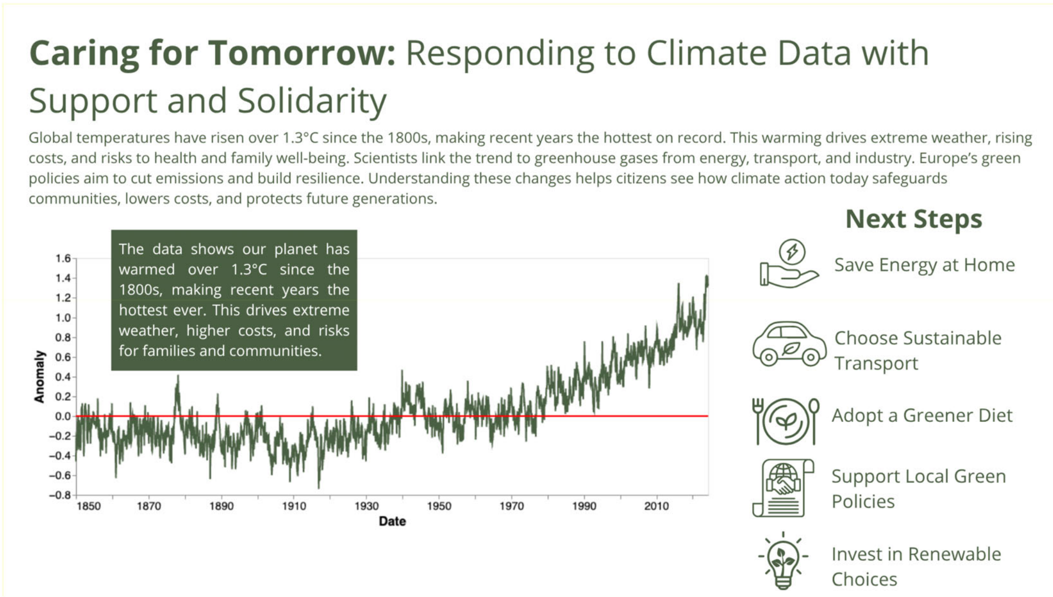

FIGURE 4 The final data story for general audience.

AI-DIKW framework to adapt narratives dynamically, ensuring relevance and engagement.

6 | CONCLUSIONS AND FUTURE WORK

This paper has described a possible strategy for using Generative AI in data storytelling as a co-designer for data stories. In particular, the AI-DIKW framework has been presented, a model based on the S-DIKW framework to transform data into a story. In the transition from one level of the framework to another, Generative AI has been used in three ways: as an assistant for extracting insight from data, as an assistant to delve into the story topic, to extract relevant context, and as a suggester of possible implementation strategies for future steps to follow after reading the story.

Using Generative AI in data storytelling can potentially raise ethical issues (Al-Kfairy et al., 2024; Stahl & Eke, 2024). When using Generative AI in data storytelling, at least two primary ethical issues should be considered: bias and misinformation (Maleki et al., 2024). Firstly, bias in AI refers to systemic and unjustified preferences, stereotypes, or prejudices in AI systems due to altered training data (Roselli et al., 2019). This can result in narratives that inadvertently perpetuate stereotypes or unfair representations of certain groups, undermining the objectivity and fairness of data stories. Secondly, misinformation can arise when AI systems generate plausible-sounding content that does not correspond with reality. This may lead to disseminating misleading information and building fake data stories that seem plausible. To overcome these ethical issues, the approach followed in this paper was to review the content produced by Generative AI tools.

A validation process is crucial to ensure the effectiveness of AI-driven data storytelling. A robust evaluation process would involve direct feedback from real-world audiences to further validate and refine the effectiveness of AI-generated data stories. An alternative solution to validate stories is using Generative AI. Cosentino et al. have used Generative AI as an evaluation tool in review classification (Cosentino et al., 2024). A preliminary study (Lo Duca & Yocco, 2025) has already demonstrated that Generative AI can serve as a useful tool for an initial validation of stories when direct access to the target audience is not immediately available. That study compared human and AI evaluations, showing that AI can approximate the characteristics of a specific audience, with the exception of emotional responses. Nevertheless, the field of evaluating AI outputs is still in rapid development. Multiple techniques are currently being explored, but none has yet emerged as a definitive standard. Two promising approaches are *rubric-based evaluation*, in which predefined evaluation criteria are established and then applied by the Generative AI itself to its generated outputs, and *grounding-based evaluation*, in which the AI assesses the quality of its output against a reference solution (Lanham, 2025). Both techniques offer structured ways to assess the reliability and quality of AI-produced results, but their robustness and transferability to real-world data storytelling scenarios remain to be further tested. Future work will therefore focus on consolidating more robust evaluation strategies, particularly leveraging multi-agent systems, to better simulate audience diversity and to strengthen the reliability of AI-assisted validation. In future research, new metrics will be defined to fully evaluate the generated stories' impact on specific audiences, such as their change in behavior after encountering the story. To obtain qualitative, relevant results, a multidisciplinary approach should be followed, including experts in data, communication, and psychology.

A further development of the research in this field could be the use of Generative AI to suggest stories based on characters extracted directly from the data or to try the automatic generation of stories starting from different types of data, following the path already taken in the field of archaeology (Buscemi & Lo Duca, 2024).

Finally, future work will further explore how wisdom in AI-assisted data storytelling may integrate evaluation of consequences and value-based decision-making, beyond audience-centered adaptation. The reflections point to a promising direction for expanding the AI-DIKW framework in this sense.

ACKNOWLEDGMENT

Open access publishing facilitated by Consiglio Nazionale delle Ricerche, as part of the Wiley - CRUI-CARE agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the global anomalies and index data: https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies.

ORCID

Angelica Lo Duca https://orcid.org/0000-0002-5252-6966

REFERENCES

Al-Kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *Informatics*, *11*(3), 58.

Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A survey on RAG with LLMs. *Procedia Computer Science*, *246*, 3781–3790.

Battle, L., & Ottley, A. (2023). What exactly is an insight? A literature review. In *2023 IEEE visualization and visual analytics (VIS)* (pp. 91–95). IEEE.

Berengueres, J. L., & Sandell, M. (2019). *Introduction to data visualization & storytelling*.

Brooks, I., D'Agostino, M., Marti, M., McDowell, K., Mejia, F., Betancourt-Cravioto, M., Gatzke, L., Hicks, E., Kyser, R., Leicht, K., Dos Pereira Santos, E., Saw, J. J., Tomio, A., & Garcia Saiso, S. (2023). An anti-infodemic virtual center for the americas. *Revista Panamericana de Salud Pública/Pan American Journal of Public Health*, *47*, e5. https://doi.org/10.26633/RPSP.2023.5

Bryant, J., & Mukherjee, A. (2023). *The future of finance with ChatGPT and power BI: Transform your trading, investing, and financial reporting with ChatGPT and power BI*. Packt Publishing.

Buscemi, F., & Lo Duca, A. (2024). Combining generative AI and archaeology to build data-driven stories. In *Proceedings of AIUCD*. ACM.

Cairo, A. (2012). *The functional art: An introduction to information graphics and visualization*. New Riders.

Chen, C., & Liu, Z. (2023). The state of the art in creating visualization corpora for automated chart analysis. *Computer Graphics Forum*, *42*, 449–470.

Clay, T. J., Steel, Z. J. D. C., & Jacobs, C. (2024). Human-computer interaction: A literature review of artificial intelligence and communication in healthcare. *Cureus*, *16*, 11.

Coelho, D., & Mueller, K. (2020). Infomages: Embedding data into thematic images. *Computer Graphics Forum*, *39*(3), 593–606.

Cosentino, C., Gündüz, M., Marozzo, F., & Birim, Ş. Ö. (2024). Exploiting large language models for enhanced review classification explanations through interpretable and multidimensional analysis. In *Proceedings of the 27th international conference on discovery science (DS 2024)*. Springer.

Dykes, B. (2019). *Effective data storytelling: How to drive change with data, narrative and visuals*. John Wiley & Sons.

Dykes, B. (2024). The future of data storytelling is augmented, not automated. Retrieved from https://www.forbes.com/sites/brentdykes/2024/02/27/the-future-of-data-storytelling-is-augmented-not-automated/ (last accessed 2024/09/12)

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, *22*(3), 110–161.

GitHub. (2024). Github copilot: Your ai pair programmer. Retrieved from https://github.com/features/copilot (An AI-powered code completion tool developed by GitHub and OpenAI)

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). Chatgpt is not all you need. A state of the art review of large generative ai models. arXiv preprint.

He, Y., Cao, S., Shi, Y., Chen, Q., Xu, K., & Cao, N. (2024). Leveraging large models for crafting narrative visualization: A survey. arXiv preprint arXiv:2401.14010.

Hine, D. W., Phillips, W. J., Cooksey, R., Reser, J. P., Nunn, P., Marks, A. D., Marks, A. D. G., Loi, N. M., & Watt, S. E. (2016). Preaching to different choirs: How to motivate dismissive, uncommitted, and alarmed audiences to adapt to climate change? *Global Environmental Change*, *36*, 1–11.

Horn, R. E. (1998). *Visual language: Global communication for the 21st century*. MacroVU Inc.

Huang, Y., Chen, Y., Zhang, H., Li, K., Fang, M., Yang, L., Li, X., Shang, L., Xu, S., Hao, J., Shao, K., & Wang, J. (2025). Deep research agents: A systematic examination and roadmap. arXiv. https://arxiv.org/abs/2506.18096

Klein, G. (2013). *Seeing what others don't: The remarkable ways we gain insights*. Public Affairs.

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Lanham, M. (2025). *AI agents in action*. Manning Publications.

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

Li, H., Wang, Y., Liao, Q. V., & Qu, H. (2023). Why is AI not a panacea for data workers? An interview study on human-ai collaboration in data storytelling. arXiv preprint.

Li, H., Wang, Y., & Qu, H. (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). ACM.

Li, H., Ying, L., Shen, L., Wang, Y., Wu, Y., & Qu, H. (2025). Composing data stories with meta relations. arXiv preprint arXiv: 2501.03603.

Liew, A., & Mueller, K. (2022). Using large language models to generate engaging captions for data visualizations. arXiv preprint arXiv:2212.14047.

Liu, C., Guo, Y., & Yuan, X. (2023). Autotitle: An interactive title generator for visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *30*, 5276–5288.

Lo Duca, A. (2023). Towards a framework for AI-assisted data storytelling. In *International conference on web information systems and technologies, WEBIST-proceedings*.

Lo Duca, A. (2024a). *Data storytelling with Altair and AI*. Manning Publications.

Lo Duca, A. (2024b). Using retrieval augmented generation to build the context for data-driven stories. In *Proceedings of the 19th international joint conference on computer vision, imaging and computer graphics theory and applications-volume 1: Ivapp* (pp. 690–696). SciTePress.

Lo Duca, A. (2025). *Become a great data storyteller. Learn how you can drive change with data*. John Wiley and Sons.

Lo Duca, A., & McDowell, K. (2025). Using the S-DIKW framework to transform data visualization into data storytelling. *Journal of the Association for Information Science and Technology*, *76*(5), 803–818. https://doi.org/10.1002/asi.24973

Lo Duca, A., & Yocco, V. (2025). Exploring the use of generative AI for assessing data-driven stories. In *Proceedings of 9th international conference on computer-human interaction research and applications (CHIRA)*. Springer.

Longnecker, N. (2023). Good science communication considers the audience. In *Teaching science students to communicate: A practical guide* (pp. 21–30). Springer.

Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI hallucinations: A misnomer worth clarifying. In *2024 IEEE conference on artificial intelligence (CAI)* (pp. 133–138). IEEE.

Matei, S. A., & Hunter, L. (2021). Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism. *Information Society*, *37*(5), 312–322. https://doi.org/10.1080/01972243.2021.1951415

Matuk, C., Amato, A., Davidesco, I., Rubel, L., Stornaiuolo, A., Bumbacher, E., & Woods, P. J. (2022). Data storytelling in the classroom. In *Proceedings of the 16th international conference of the learning sciences-ICLS 2022* (pp. 1779–1786). International Society of the Learning Sciences.

McDowell, K. (2021). Storytelling wisdom: Story, information, and DIKW. *Journal of the Association for Information Science and Technology*, *72*(10), 1223–1233. https://doi.org/10.1002/asi.24466

Milesi, M. E., & Martinez-Maldonado, R. (2024). Data storytelling in learning analytics? A qualitative investigation into educators' perceptions of benefits and risks. In *Proceedings of the 14th learning analytics and knowledge conference* (pp. 167–177). ACM.

NOAA. (2024). The global anomalies and index data. Retrieved from https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies (last accessed 2024, September 17)

Ojo, A., & Heravi, B. (2018). Patterns in award winning data storytelling: Story types, enabling tools and competences. *Digital Journalism*, *6*(6), 693–718.

OpenAI. (2024a). Chatgpt. Retrieved from https://openai.com/chatgpt/ (last accessed 2024/09/13)

OpenAI. (2024b). Dall-e 3. Retrieved from https://openai.com/index/dall-e-3/ (last accessed 2024/09/13)

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The persuasive power of data visualization. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2211–2220.

Pasquinelli, M., & Joler, V. (2021). The nooscope manifested: AI as instrument of knowledge extractivism. *Ai & Society*, *36*(4), 1263–1280.

Renda, G., Daquino, M., & Presutti, V. (2023). Melody: A platform for linked open data visualisation and curated storytelling. In *Proceedings of the 34th ACM conference on hypertext and social media* (pp. 1–8). Association for Computing Machinery.

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. In *Companion proceedings of the 2019 world wide web conference* (pp. 539–544). ACM.

Sarica, H. C., & Yildirim, D. (2024). Data storytelling in education: A systematic review. In *INTED2024 proceedings* (pp. 6990–6997). IATED Academy.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Shao, H., Martinez-Maldonado, R., Echeverria, V., Yan, L., & Gasevic, D. (2024). Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension? In *Proceedings of the 2024 CHI conference on human factors in computing systems (CHI'24)* (pp. 1–21). Association for Computing Machinery.

Singer, J. B. (2008). Five Ws and an H: Digital challenges in newspaper newsrooms and boardrooms. *International Journal on Media Management*, *10*(3), 122–129. https://doi.org/10.1080/14241270802262468

Stahl, B. C., & Eke, D. (2024). The ethics of chatgpt–exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*, 102700.

Van Meter, H. J. (2020). Revising the dikw pyramid and the real relationship between data, information, knowledge, and wisdom. *Law, Technology and Humans*, *2*(2), 69–80.

Wang, Q., Chen, Z., Wang, Y., & Qu, H. (2021). A survey on ml4vis: Applying machine learning advances to data visualization. *IEEE Transactions on Visualization and Computer Graphics*, *28*(12), 5134–5153.

Wang, X., Wu, Z., Huang, W., Wei, Y., Huang, Z., Xu, M., & Chen, W. (2023). Vis+ ai: Integrating visualization with artificial intelligence for efficient data analysis. *Frontiers of Computer Science*, *17*(6), 176709.

Wei, Z., Qu, H., & Xu, X. (2025). Telling data stories with the hero's journey: Design guidance for creating data videos. *IEEE Transactions on Visualization and Computer Graphics*, *31*(1), 962–972. https://doi.org/10.1109/TVCG.2024.3456330

Yang, L., Xu, X., Lan, X. Y., Liu, Z., Guo, S., Shi, Y., Qu, H., & Cao, N. (2022). A design space for applying the Freytag's pyramid structure to data stories. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 922–932. https://doi.org/10.1109/TVCG.2021.3114774

Yu, F., & Stasko, J. (2024). More than data stories: Broadening the role of visualization in contemporary journalism. *IEEE Transactions on Visualization and Computer Graphics*, *30*(8), 5240–5259. https://doi.org/10.1109/TVCG.2023.3287585

Yuan, J., Chen, C., Yang, W., Liu, M., Xia, J., & Liu, S. (2021). A survey of visual analytics techniques for machine learning. *Computational Visual Media*, *7*, 3–36.

Zapata-Rivera, D., Torre, I., Lee, C.-S., Sarasa-Cabezuelo, A., Ghergulescu, I., & Libbrecht, P. (2024). Editorial: Generative AI in education. *Frontiers in Artificial Intelligence*, *7*, 1532896. https://doi.org/10.3389/frai.2024.1532896

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lo Duca, A. (2025). Using generative AI to co-design data-driven stories. *Journal of the Association for Information Science and Technology*, *76*(13), 1786–1802. https://doi.org/10.1002/asi.70036