

InSituTale: Enhancing Augmented Data Storytelling with Physical Objects

Kentaro Takahira

The Hong Kong University of
Science and Technology
Hong Kong, China
ktakahira@connect.ust.hk

Yue Yu

The Hong Kong University of
Science and Technology
Hong Kong, China
yue.yu@connect.ust.hk

Takanori Fujiwara

Linköping University
Linköping, Sweden
University of Arizona
Tucson, USA
tfujiwara@ucdavis.edu

Ryo Suzuki

University of Colorado Boulder
Boulder, USA
ryo.suzuki@colorado.edu

Huamin Qu

The Hong Kong University of
Science and Technology
Hong Kong, China
huamin@cse.ust.hk



Figure 1: InSituTale: We developed *InSituTale*, an augmented physical data storytelling prototype to enable presenters to control visualizations through physical object manipulations, achieving coordination of physical and digital elements.

Abstract

Augmented data storytelling enhances narrative delivery by integrating visualizations with physical environments and presenter actions. Existing systems predominantly rely on body gestures or speech to control visualizations, leaving interactions with physical objects largely underexplored. We introduce *augmented physical data storytelling*, an approach enabling presenters to manipulate visualizations through physical object interactions. To inform this approach, we first conducted a survey of data-driven presentations to identify common visualization commands. We then conducted workshops with nine HCI/VIS researchers to collect mappings between physical manipulations and these commands. Guided by these insights, we developed *InSituTale*,

a prototype that combines object tracking via a depth camera with Vision-LLM for detecting real-world events. Through physical manipulations, presenters can dynamically execute various visualization commands, delivering cohesive data storytelling experiences that blend physical and digital elements. A user study with 12 participants demonstrated that *InSituTale* enables intuitive interactions, offers high utility, and facilitates an engaging presentation experience.

CCS Concepts

• Human-centered computing → Visualization design and evaluation methods.

Keywords

Visualization; Data-Driven Storytelling; Tangible Interaction; Augmented Reality; Augmented Presentation; Video

ACM Reference Format:

Kentaro Takahira, Yue Yu, Takanori Fujiwara, Ryo Suzuki, and Huamin Qu. 2025. *InSituTale: Enhancing Augmented Data Storytelling with Physical Objects*. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*, September 28–October 1, 2025, Busan, Republic of Korea. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3746059.3747678>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UIST '25, Busan, Republic of Korea

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-2037-6/2025/09 <https://doi.org/10.1145/3746059.3747678>

1 Introduction

Augmented data storytelling enriches data-driven narratives by integrating visualizations with physical environments and presenter actions. Recent systems have demonstrated the potential of controlling visualizations through body gestures [15, 22, 29, 53] and speech [32, 57], enhancing performative aspects and enabling improvisational storytelling experiences that respond dynamically to audience engagement. However, these systems predominantly focus on gesture- and speech-driven interactions, overlooking the role of physical objects as integral storytelling components or interaction mediums.

This limited focus represents a notable gap. In real-world data storytelling, such as product demonstrations or educational contexts, physical objects often serve as referents for visualizations, playing a central role in conveying data narratives [16, 28, 70]. The coordinated manipulation of physical objects and associated visualizations can enhance audience understanding and engagement, as illustrated in professionally edited augmented videos [25, 50]. Despite the rich affordances of physical objects, including grasping, rotating, combining, and transforming, existing augmented data storytelling systems offer limited support for these interactions [22, 29, 32, 32, 33]. Consequently, current solutions often fail to deliver cohesive storytelling experiences that effectively blend physical and digital elements.

In this paper, we introduce *augmented physical data storytelling*, a novel approach enabling presenters to intuitively control visualizations through direct manipulations of physical objects. We expect that coupling physical object interactions meaningfully with visualization responses can offer presenters intuitive control and expressive storytelling capabilities, seamlessly blending physical and digital elements. Additionally, tangible interactions leveraging rich physical affordances are expected to reduce presenters' cognitive load and offer more expressive interactions than purely gesture-based methods [60].

To inform our approach, we first conducted a survey of data-driven presentations to identify a relevant set of visualization commands (e.g., showing charts, selecting data points, and changing chart types). Subsequently, we held workshops with nine HCI/VIS researchers to identify intuitive mappings between physical manipulations and these visualization commands. The workshops produced diverse user-generated mappings categorized into six groups: 1) appearance-based, 2) movement-based, 3) arrangement-based, 4) gesture-based, 5) affordance-based, and 6) visualization-based interactions. We analyzed these ideas and identified common mappings for each command. Drawing upon these insights, we iteratively developed *InSituTale*, a prototype supporting augmented physical data storytelling. This process also yielded five critical design considerations: 1) physical space interaction detection, 2) dynamic visualization placement, 3) minimization of interaction ambiguity, 4) smooth storytelling flow, and 5) context-aware presenter guidance.

InSituTale consists of two primary modes: presentation and authoring. In the presentation mode, presenters control visualizations in real time using physical object manipulations. A depth camera captures physical manipulations, including pointing gestures, lifting objects, adjusting distances between objects, and moving objects closer to or farther from the camera. Additionally, a vision-language model supports real-time detection of custom-defined changes in object states (e.g., “Is the banana peeled?”), which can trigger corresponding visualization updates. These capabilities enable presenters to execute a wide range of

visualization commands, including showing/hiding charts, scaling, composing or decomposing charts, selecting individual data points or series, switching chart types, and transitioning between overview and detailed views. In the authoring mode, presenters configure scenes by assigning visualizations to physical objects, adding annotations, and defining customizable commands.

We evaluated *InSituTale* with 12 participants, assessing its usability, utility, and learnability. Quantitative and qualitative results confirmed that *InSituTale* supports intuitive interactions, provides high utility, and facilitates engaging presentation experiences. This paper makes the following main contributions:

- ◊ **Novel Concept and Interaction Design:** We introduce *augmented physical data storytelling* and present insights into intuitive mappings between physical object manipulations and visualization commands through design workshops.
- ◊ **System Design and Implementation:** We developed *InSituTale*, a prototype that integrates object tracking and a vision-language model to support real-time, physically driven interactions with visualizations.
- ◊ **User Study:** We report on user studies highlighting system strengths, limitations, and implications for designing future augmented data storytelling systems.

2 Related work

Our research examines the intersection of data-driven storytelling, augmented presentation, and physical object-based interactions for data visualization. We review these domains and illustrate how our work complements prior studies.

2.1 Data-Driven Storytelling

Data-driven storytelling [5, 30], where presenters deliver real-time, data-driven narratives, is a valuable practice in various contexts, such as organizational decision-making and public communication [5]. In particular, improvisational storytelling, characterized by small audiences and spontaneous data communication, known as *jam-session-style* [5], requires presenters to adapt their narratives in response to audience interactions, creating personalized and engaging presentation experiences by interacting with visualizations. Systems designed to support interaction in data-driven storytelling have been proposed, from AR-based [15, 22, 60] to screen-based presentations [5, 30, 57]. These systems focus on facilitating easy and direct manipulation of visualizations rather than at the slide level [5, 30]. Also, they often utilize gestures or speech to enhance the performative quality of the presentations [22, 57, 60].

In jam-session-style scenarios, physical objects can play a pivotal role by enriching the storytelling experience [9, 44]. As props, these objects serve as the central focus of the presentation and tangible anchors for abstract data, helping bridge the gap between complex information and audience comprehension. For instance, product demonstrations often pair real physical products with visualizations displayed on separate screens or papers [28]. Similarly, Hans Rosling's iconic presentations showcased the powerful use of physical objects like boxes and stones, which he manipulated by stacking, hiding, and relocating to illustrate demographic dynamics [46, 49, 51, 52]. These techniques contextualize abstract data with physical objects, making the information more tangible and understandable while also boosting audience engagement. The theatrical theory further underscores the performative quality of props, highlighting their ability to captivate audiences and strengthen communication [8]. Despite

these compelling examples, the use of physical objects in data storytelling remains largely overlooked [44, 60]. In this paper, we fill this gap by introducing augmented physical data storytelling.

2.2 Augmented Presentation

Augmented presentations [32, 53], which overlay digital content onto presenters and their physical environment, are gaining popularity across domains such as education [32, 45], advertising [61], and business presentation [22]. Existing research has explored various approaches to augmenting presentations [6, 10, 22, 29, 32, 33, 36, 42, 45, 53]. Saquib *et al.* [53] introduced body-driven graphics that map visualizations to specific body parts, dynamically adapting to presenters' movements. Hall *et al.* [22] developed Augmented Chironomia, a system that enables gesture-based visualization control for remote presentations. RealityTalk [32] employs a keyword-matching system to recognize spoken words and generate corresponding graphic elements that presenters can manipulate through gestures. Liu *et al.* [36] proposed PoseTween, an authoring tool that animates objects based on human poses, ensuring natural coordination between human actions and object animations. Elastica [6] addresses challenges in recognition errors and presenter mistakes by allowing dynamic adjustments to predefined graphic animations through speech and gestures. RealityEffects [33] augments volumetric 3D scenes, enabling users to bind captured physical objects with annotated visual effects that dynamically respond to physical motion.

While these approaches offer various modalities in different ways, they offer limited coordination between physical objects and visualizations. Most existing systems either overlook the role of physical objects [6, 22, 36, 53] or focus on basic interactions, such as tracking objects and having visuals follow their movements [32, 33]. Consequently, they fail to fully utilize the diverse manipulations that physical objects afford, such as stacking, rotating, colliding, and bringing multiple objects closer together. By contrast, traditional video-editing tools, while capable of incorporating physical props, lack the real-time flexibility required for improvisational storytelling, particularly in mid- to small-scale settings [5]. Our work addresses this gap by designing augmented presentations that facilitate effective coordination between physical objects and visualizations, with a particular focus on data storytelling.

2.3 Interacting with Visualization Using Physical Objects

Interacting with visualizations through physical objects offers substantial benefits, including reduced cognitive load and learning cost [27, 56]. Physical objects provide an intuitive and direct way to engage with data, leveraging users' natural spatial reasoning and motor skills to minimize the cognitive effort required for interaction [63]. To further bridge the gap between the virtual and physical realms, embedded data representations integrate visualizations with the physical objects or spaces to which it refers, known as referents [12, 13, 68]. Techniques for integrating visualizations with physical referents include adding labels [7, 34], overlaying information [37, 69], nearby placement [11, 67], and spatial projections [41]. These methods create a seamless connection between data and its physical referents.

When physical objects are used as referents, their properties—such as being pickable, stackable, or combinable—together with their semantic relationships to visualizations, open up diverse and meaningful opportunities for interacting with visualizations [54,

62, 64]. For instance, Uplift [14] allows users to display building information by picking up the corresponding scale models from a table. Satriadi *et al.* [55] explored tangible globes for geospatial visualizations, enabling physical manipulation of embedded maps and data. Additionally, Active Proxy Dashboard [54] uses scale models to interact with dashboards, supporting operations like filtering data by picking up specific models or authoring composite visualizations by bringing multiple models together. Herman *et al.* [24] introduced landscape models that integrate geospatial data and simulations, allowing users to adjust cutting planes and simulations precisely within the same spatial coordinate system as the physical model.

While these studies provide valuable insights into interaction design with physical objects in augmented environments, most focus on analytics purposes, where the primary goal is to support data exploration [13]. In contrast, data storytelling introduces unique requirements, such as the performative aspects of interaction [22], real-time responsiveness to audience discussions [5], and intuitive, low-cost interactions that allow presenters to focus on delivering their narrative without errors [6, 30]. Studies on data physicalization further highlight the communicative potential of tangible artifacts to make abstract data more tangible and engaging [3, 18, 59]. However, many of these approaches remain static, constrained by technical limitations that prevent dynamic visualization changes or real-time interactions—features critical for data storytelling. Our work contributes to this field by designing an interaction framework for real-time data storytelling, leveraging physical objects as referents.

3 Augmented Physical Data Storytelling

This section introduces the concept of *augmented physical data storytelling* along with its key features.

3.1 Concept

We propose *augmented physical data storytelling*. An approach that integrates physical object manipulations with data visualizations to deliver engaging, real-time narratives. This approach introduces an interface that synchronizes the movement and transformation of physical objects with corresponding changes in visualizations, blending physical artifacts, digital visuals, and presenter gestures into a cohesive mixed-reality experience. Unlike prior methods that rely on extensive post-production, our approach enables seamless, live presentation without the need for video editing. Furthermore, unlike previous studies that use physical objects for visualization interactions, our approach is designed to support novel interaction techniques tailored to real-time data storytelling contexts.

Our approach draws inspiration from the presentation style of Hans Rosling, who captivated audiences by integrating multiple modalities—gestures, speech, physical objects, and visuals—to create an engaging and cohesive storytelling experience. Rosling's presentations often involved post-edited animations on visualizations and synchronized them with his gestures to simulate direct control over the visuals [47, 48, 50]. He also used tangible props like boxes, paper rolls, and stones to make abstract demographic data more relatable [49, 51, 52].

Inspired by this interplay between the physical and the visual, our goal is to realize a real-time, end-to-end system that preserves this expressiveness while enabling real-time storytelling. We envision applications across domains such as education, product demonstrations, and public presentations, where physical objects

serve as essential narrative props, and their coordination with visualizations enhances the clarity and communicative power of data-driven narratives.

3.2 Key Features

To realize the concept of *augmented physical data storytelling*, the system incorporates the following core features:

Coordination of Physical Objects and Visualizations. The system enables real-time synchronization between physical object manipulations and corresponding changes in visualizations. Movements such as repositioning, arranging, or transforming objects dynamically trigger visual updates, fostering a seamless and intuitive link between the physical and digital elements. Our approach supports a range of semantic couplings, from loosely performative to tightly meaningful. For example, lifting a physical object may simply serve to reveal a related visualization in a performative manner. In contrast, changing the state of a physical object, such as peeling a banana, can trigger visual updates that reflect semantic meaning, such as nutritional information tied to the object's edible state.

Support for Diverse Interactions. To meet the expressive needs of data storytelling, the system supports a wide range of interaction types with visualizations [22, 30, 60]. This flexibility enables presenters to construct rich narratives that adapt to varying data complexities and audience engagement.

Improvisational Presentation. The system is designed to support non-linear, adaptive storytelling, as emphasized in prior work [5]. Presenters can adjust the narrative flow in real time based on audience reactions or spontaneous insights, enabling more engaging and personalized presentations.

4 Design Workshops to Solicit Interactions

Through a formative workshop, we explored how physical object manipulations can naturally coordinate with visualization commands used in data storytelling. For instance, when introducing a new product, a presenter may pick it up, point to it, move it closer to the audience, or adjust its shape to emphasize particular features. Suppose such physical manipulations can be meaningfully mapped to visualization commands through well-designed interaction mechanisms. In that case, it raises the question of which mappings between physical object manipulations and visualization commands are intuitive and effective for storytelling. To answer this question, we conducted design workshops to identify effective mappings that could inform the design of augmented physical data storytelling systems.

4.1 Visualization Commands for Storytelling

Our first goal was to identify a relevant set of visualization commands for data storytelling, such as showing, hiding, and scaling visualizations. Given the wide range of possible commands, we focused on those most relevant to real-time presentation settings. To ensure that the selected commands reflect authentic storytelling practices, we conducted an exploratory survey of publicly available videos showcasing augmented or interactive presentations with data visualizations. We performed keyword-based searches on platforms such as YouTube and Vimeo using terms including “data-driven storytelling,” “AR presentation,” “virtual presentation,” and “data presentation.” Due to the absence of standardized repositories or consistent keywords, we adopted a manual, iterative collection strategy. Rather than starting from a

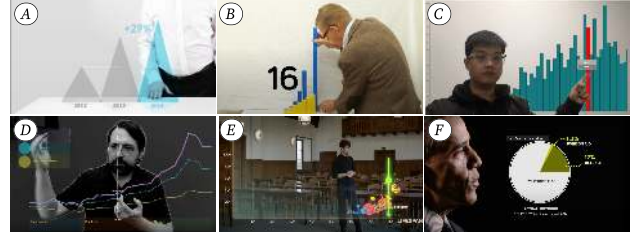


Figure 2: Interactions in Data-Driven Presentations: (A) scale up [1], (B) overlay bar charts [17], (C) select a bar [29] (D) select data series [26], (E) change a bubble chart to a bar chart [58], (F) show a detailed view [66].

broader pool, we directly collected matching examples based on their relevance to our goals. We selected videos that featured a presenter delivering data-driven content to an audience, using visualizations in conjunction with physical or gestural interactions. This process resulted in a curated dataset of 31 videos (see the supplemental materials). While not exhaustive, this dataset served as a resource for identifying common visualization commands used in practice. Three researchers independently analyzed this dataset and then collaboratively discussed their observations to reach a consensus. Through this process, we distilled a set of key visualization commands, including:

Show/Hide: Toggling the visibility of a visualization during the presentation. This command appeared in nearly all the videos, as it serves as a fundamental mechanism for controlling the narrative flow.

Scale: Adjusting the size of a visualization to control focus and legibility. This command was commonly used to draw the audience's attention to a specific chart or element, especially when emphasizing the details (Fig. 2-A).

Compose/Decompose: Merging multiple visualizations into a unified view or separating them into distinct parts. For instance, overlaying two bar charts can help compare two products. This technique was often used to emphasize part-whole relationships or contrast different data sources (Fig. 2-B).

Select/Deselect Data Points: Highlighting an individual data point. This command was often used to draw attention to a specific value during narration (Fig. 2-C).

Select/Deselect Data Series: Selecting a group of data points that share the same category (e.g., all countries in Asia). This was commonly used to highlight patterns or differences across categories (Fig. 2-D).

Change Chart Types: Switching between chart types (e.g., from a bar chart to a line chart) to convey different aspects of the same data source. This command was used to show multiple perspectives on the data (Fig. 2-E).

Change Data Sources: Switching data sources being visualized to align with the narrative context or the state of the physical object. This can be done alongside a change in the visualization type.

Hierarchical Navigation: Navigating between overview and detail views within a visualization. This commonly followed a “zoom-in” narrative structure—for example, presenting a broad category using a pie chart and then revealing its breakdown into subcategories (Fig. 2-F).

4.2 Selection of Physical Objects

To ground discussions during the workshop, we prepared a set of physical objects to explore interaction mappings. Guided by prior work on object affordances [20], we focused on three key dimensions that likely affect users’ physical object manipulations: size, grabbability, and edge characteristics. We selected tabletop-sized objects to reflect typical use contexts—such as product demonstrations—and to support diverse hand-scale interactions. Based on these considerations, we selected a set of objects spanning a range of dimensions, including a cup, wine bottle, banana, toy car, backpack, and laptop (Table 1).

Table 1: Selected Physical Objects: We provided the six objects to cover diverse ranges of physical properties.

	Cup	Bottle	Banana	Toy Car	Backpack	Laptop
						
Size	S	M	M	S	L	M
Grabbable	✓	✓	✓	✓	✗	✗
Curved Edge	✓	✓	✓	✗	✓	✗

4.3 Ideation Workshop

The ideation workshop involved researchers in structured activities, where they engaged in brainstorming and group discussions to develop a variety of interaction ideas. Three sessions were held, each with three participants. Sessions continued until the proposed mappings reached saturation; while saturation wasn’t formally measured, 90% of the final participant’s mappings in the last workshop overlapped with earlier ones.

4.3.1 Participants. Our workshop involved nine participants (W1–W9), including researchers specializing in Visualization (VIS) and Human-Computer Interaction (HCI). Participants (four HCI, five VIS) had 4–10 years of experience and selected for their expertise in interaction design, following methodologies in similar studies [23, 60, 62]. The participants ranged in age from 23 to 32 and included seven males and two females.

4.3.2 Materials. Participants received a set of visualization commands with accompanying illustrations to support comprehension (see supplemental materials). These materials remained accessible throughout the workshop. Participants were also provided with the six selected physical objects and encouraged to experiment with potential user manipulations for the given commands. To facilitate idea generation, we introduced several example mappings, which served as a priming technique [39], commonly used in previous studies [23, 60, 62], to stimulate creative thinking and clarify possible interaction mappings.

4.3.3 Procedure. Each 70-minute session comprised three phases: briefing, individual brainstorming, and group discussion. In the briefing, participants completed a consent form, followed by a 10-minute introduction covering the research background and workshop objectives. We explained the visualization commands and demonstrated the example mappings while emphasizing that these were merely examples and encouraging creative thinking beyond them. During individual brainstorming (25–30 minutes), participants explored ways to perform each visualization command by manipulating the provided physical objects. The participants were asked to record their mapping ideas on worksheets.

They were also encouraged to think freely without concern for technical limitations. In the group discussion (20–30 minutes), participants demonstrated their ideas using the physical objects, discussed possible extensions, and explained the reasoning behind each mapping. The entire discussion was recorded and documented in the instructor’s notes for later analysis.

4.3.4 Data Analysis. Two authors organized the workshop outcomes, identifying a total of 143 unique manipulations across the six physical objects. Proposed manipulations were initially coded using a reference set derived from prior studies [9, 20, 43], which included common manipulations such as tapping, colliding, and stacking. An open-coding approach was applied to manipulations beyond this initial set. Coding was refined iteratively through discussions among two authors until consensus was achieved.

4.4 Results

We summarized the results in terms of prominent physical manipulation categories and frequently proposed mappings.

4.4.1 Manipulation Categories. We synthesized the diverse manipulation ideas from participants into 15 distinct types, organized into six overarching categories based on the nature of the manipulations and the intentions.

Ap Appearance-based Interactions control what is exposed to the system or the audience. This includes *visibility control* (e.g., showing or hiding a physical object from the camera’s view), *appearance control* (e.g., transforming the shape of a physical object), and *access control* (e.g., revealing or concealing the contents when opening a bag or removing a bottle cap).

M Movement-based Interactions involve changing a physical object’s position or orientation. The proposed manipulations include *lifting*, *relocating*, *changing orientation*, or *changing the distance from the camera*.

Ar Arrangement-based Interactions relate to the spatial arrangement of multiple physical objects. This category includes *adjusting distances among physical objects* and *isolating a specific object* visually or spatially from others (e.g., lifting one while leaving others in place).

G Gesture-based Interactions are hand gestures performed on or near an object, including *tapping*, *shaking*, *pinching*, and *drawing* on surfaces with a finger.

Af Affordance-based Interactions utilize the inherent affordances of a physical object to trigger commands. Examples include moving a toy car door, tightening a bag strap, drinking a bottle of wine, or pressing a laptop’s spacebar.

V Visualization-based Interactions directly control visualizations without involving physical objects. Participants typically proposed these manipulations when perceived to be more immediate or intuitive than controlling visualizations through physical objects. Common examples include *pointing at visuals* and *enclosing them with both hands*.

4.4.2 Dominant Mappings. Based on the categorized types of manipulations, we analyzed frequently proposed mappings between physical manipulations and visualization commands (Fig. 3). Refer to our supplementary materials for object-specific analyses.

Show/Hide: Participants frequently mapped this command to **Ap Appearance-based** and **M Movement-based Interactions**, such as **Ap** *controlling visibility from the camera* (35%), **M** *lifting the object* (27%), and **Ap** *controlling access* (26%). Controlling visibility or lifting was commonly suggested for small, easily

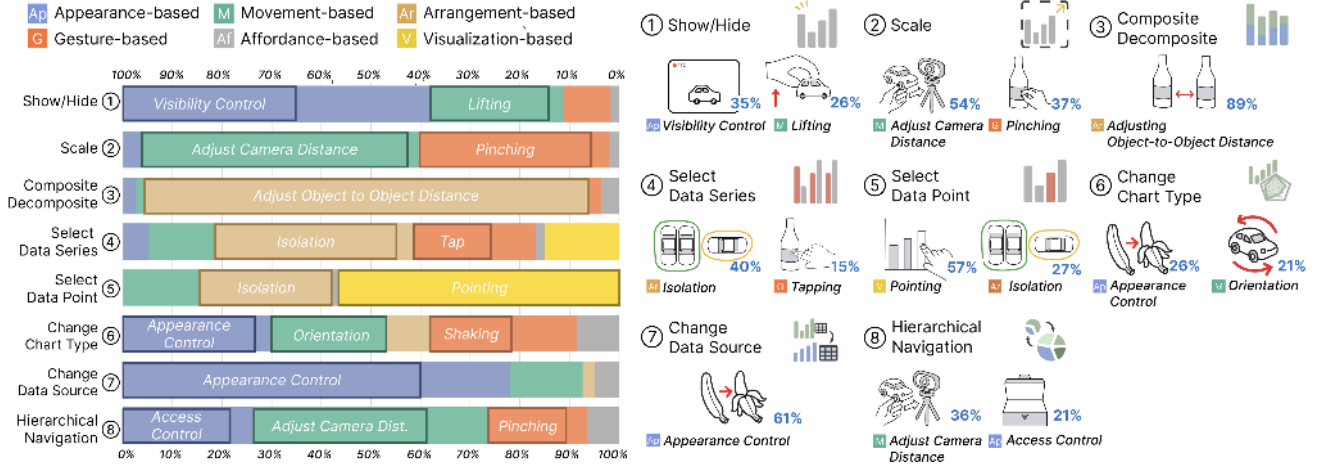


Figure 3: Workshop Results: This illustrates mappings between physical manipulations and visualization commands across various physical objects. 15 types of manipulations are grouped into six categories. For each visualization command, we present the cumulative percentage of manipulations within each category, along with the top two most frequently proposed manipulations across categories. A detailed breakdown of individual objects is provided in the appendix.

graspable objects, while access controls were favored for heavier or less grabbable items (e.g., opening a backpack).

Scale: This command was primarily associated with **Movement-based** and **Gesture-based Interactions**, particularly **M** adjusting the object distance from the camera (54%) and **G** pinching (37%). Participants often aimed to match the perceived size of the physical object from audiences with the intended scale. For larger objects with broad surfaces (e.g., laptops), pinching gestures were more common, reflecting participants' prior experiences with touchscreen scaling.

Compose/Decompose: Most participants mapped this command to **Arrangement-based Interactions**, particularly **Ar** adjusting the distance between objects (89%). The proposed gestures were consistent across all physical objects, intuitively associating the close proximity of two physical objects with the composition and the distant proximity with the decomposition.

Select/Deselect Data Points: Participants favored **Visualization-based Interactions**, especially **V** pointing gestures (57%). As W3 noted, "When the visualization is directly in front of me, it feels inefficient to involve a physical object for selecting a single data point." Others (e.g., W1) proposed **Ar** isolation (27%) for small visual targets, such as points in a scatterplot, noting that selecting tiny elements with fingers can be difficult.

Select/Deselect Data Series: **Arrangement-based Interactions**, specifically **Ar** isolation of the target object (40%), was the most favored approach. Isolation involves separating one object spatially from others to select its linked data series. As W1 noted, "When an object stands apart visually, it makes sense that the audience would focus on it—and the system could then highlight the associated data."

Change Chart Types: No dominant strategy emerged from the participants' responses. Participants proposed **Ap** changing the object's appearance (26%), such as peeling a banana or swapping it with a different-colored object, to signal a change in chart type visually. Others suggested **M** rotating the object (21%) to switch between visualizations, mapping the physical object's orientation to different chart types.

Change Data Source: Most participants associated this with **Appearance-based Interactions**, particularly **Ap**

modifying the object's appearance (61%). Similar to changing visualization type, these manipulations symbolically conveyed a shift in the underlying data source, such as replacing or revealing new aspects of the object.

Hierarchical Navigation: No single strategy dominated. Participants frequently proposed **M** adjusting the object's distance from the camera (36%), **Ap** controlling access (21%), and **G** pinching (20%). While conceptually similar to visual scaling, controlling access was often preferred for its stronger semantic alignment with the idea of revealing more detailed content. As W8 described, "Opening a backpack to show what's inside just feels like drilling down into the data."

5 Design Process

We developed *InSituTale* iteratively, drawing our workshop's results on the natural mappings between physical manipulations and visualization commands from our workshop. Throughout this iterative design process, we identified five critical design considerations for augmented physical data storytelling:

D1 Physical Space Interaction Detection: Tracking physical objects and presenter gestures in 3D physical space to ensure reliable interaction detection.

D2 Dynamic Visualization Placement: Automatically positioning visualizations to prevent them from occluding a presenter and physical objects.

D3 Minimization of Interaction Ambiguity: Providing robust mappings and a minimal set of commands tailored to each storytelling scenario to prevent unintended interactions.

D4 Smooth Storytelling Flow: Balancing manual and automated visualization controls to support improvisational presentation while minimizing repetitive command execution that may disrupt storytelling flow.

D5 Context-Aware Presenter Guidance: Providing real-time, scene-specific cues to help presenters remember available interactions and maintain narrative coherence.

5.1 Iterative Design

We iteratively refined *InSituTale*'s interaction mechanisms to enhance the overall presentation experience. The prototypes



Figure 4: Presentation and Authoring Modes: *InSituTale* consists of presentation and authoring mode. (A) In presentation mode, the presenter interacts with visualizations in real time using physical objects and hand gestures, tracked by a depth camera placed in front of them. A live video stream is displayed locally for monitoring. (B) In authoring mode, the presenter configures physical objects, commands, and visualizations to be used during the presentation.

developed during the design process were tested internally by the authors and informally evaluated by external participants.

First Iteration: Supporting 3D Spatial-Aware Object Tracking. Initially, we used ArUco markers and a webcam to detect manipulations, such as moving a physical object closer to the camera or another object. Early testing showed that due to reliance on 2D screen-space tracking, these manipulations involving depth were frequently misinterpreted as horizontal or vertical screen movements, leading to inaccuracies [D1]. Additionally, the markers were often unintentionally obscured by users, causing interaction failures. These limitations motivated our shift to a markerless tracking solution with a depth camera.

Second Iteration: Improving Interaction Robustness and Flow. We integrated a depth camera with an object detection model, improving tracking accuracy and enabling reliable detection of interactions in physical space [D1]. Guided by workshop findings, we implemented common mappings between physical manipulations and visualization commands while ensuring that each manipulation remained distinct. Early user tests revealed that requiring presenters to manually execute frequently used commands, such as showing or hiding visualizations, disrupted the storytelling flow due to repetition. To mitigate this, we introduced a scene-based presentation structure, allowing presenters to assign visualizations to individual scenes and automatically control their visibility through scene transitions inspired by prior work [22] [D4]. Presenters can navigate scenes both forward and backward as needed. While this structure imposes some constraints on the overall sequence of scenes, it still offers flexibility within each scene. Presenters can vary the order of interactions and repeat actions as needed. Further testing also showed frequent misclassification of gestures, particularly between pointing and lifting. Since presenters often use their index fingers while grasping objects, these interactions were occasionally misrecognized as pointing. To mitigate this, we allowed presenters to specify a preferred pointing hand (left or right), ensuring that gestures made with the non-dominant hand would not be interpreted as pointing [D3]. Additionally, we enabled presenters to selectively activate or deactivate visualization commands on a per-scene basis to reduce the likelihood of unintended triggers [D3]. This approach reflects a balance between reducing false detections and supporting improvisation.

Third Iteration: Supporting Presenters with Contextual Awareness. We introduced support for multiple scenes, each with individually configurable commands and visualizations. Internal tests revealed that presenters frequently forgot which visualizations and interactions were associated with each scene, creating uncertainty during presentations. To overcome this issue, we added an on-screen guidance panel—visible only to presenters—to display configured interactions and associated visualizations. This panel substantially reduced cognitive load by providing immediate and contextual assistance [D5].

6 InSituTale

This section presents the final design of *InSituTale*. *InSituTale* comprises two modes: Presentation Mode (Fig. 4-A), which enables real-time storytelling through physical interactions, and Authoring Mode (Fig. 4-B), allowing presenters to configure scenes, visualizations, and interactions beforehand.

6.1 Presentation Mode

InSituTale supports real-time data storytelling through interaction with physical objects on a table. A depth camera captures the presenter’s physical space. The system generates a live video stream that combines physical footage (i.e., the presenter and objects) with augmented visualizations, viewable by both presenters and remote audiences.

6.1.1 Visualization Visibility and Scaling. Presenters can dynamically control the visibility and scale of visualizations. Placing a physical object within the camera view reveals the corresponding visualization, while hiding it can be achieved either by moving the object out of the camera’s view or by shifting it beyond a predefined distance threshold, as detected by a depth camera (Fig. 5-A1,A2). Presenters can display annotations (images or text) by pointing at a physical object with their index finger for a preset duration (Fig. 5-B1). Moving an object closer to or farther from the camera scales the visualization accordingly. Additionally, placing an object closer to the camera than a distance threshold triggers a transition to a more detailed view (Fig. 5-C1–C3).

6.1.2 Data Selection. Presenters can highlight specific data series linked to physical objects. A data series is selected by lifting



Figure 5: Supported Interactions: (A) Place an object to show a chart. (B) Point at an object using the index finger to show an annotation. (C) Move an object closer to the camera to scale the visualization and reveal details. (D) Lift an object to highlight the associated data series. (E) Point to a data point using the index finger. (F) Bring objects closer together to generate a composite chart. (G) Trigger a visualization change via a user-defined condition. Bounding boxes and the presentation panel are not shown to the audience.

an object from the table while keeping others stationary (Fig. 5-D1, D2). Individual data points can also be highlighted by pointing directly at the visualization with an index finger (Fig. 5-E1).

6.1.3 Visualization Composition and Transformation. Presenters can dynamically create composite visualizations or transform existing ones into different types. Composite visualizations are generated when multiple related objects are brought close together horizontally or vertically. For bar charts, horizontal alignment results in a clustered bar chart, while vertical alignment produces a stacked bar chart (Fig. 5-F1–F3). Other chart types—including pie, donut, radar, and line charts—are composited through visual

overlay, regardless of their spatial orientation (Fig. 5-F4). The system can also trigger visualization transformations based on detected real-world events. For instance, a presenter may define a condition such as “Is the glass filled with red wine?”. When the condition is met, the system automatically replaces the associated visualizations with pre-registered ones (Fig. 5-G1, G2).

6.1.4 Scene Control. Presenters navigate through predefined scenes, each containing unique visualizations and enabled interaction commands. They can move forward or backward between scenes as needed. Scene transitions are controlled via a mouse right-click, assuming presenters will use a commonly adopted

presentation tool, such as a ring mouse or clicker. This design choice is informed by insights from existing synchronous data storytelling systems [22], ensuring that scene transitions remain distinct from performative storytelling gestures while maintaining a smooth and controlled presentation flow.

6.1.5 Dynamic Visual Arrangement. *InSituTale* dynamically positions visualizations near their corresponding physical objects, ensuring that data remains contextually linked to the objects being referenced. When presenters move objects, visualizations update their positions in real time. The system employs an adaptive layout algorithm to prevent overlaps between visualizations, physical objects, and the presenter’s face. Composite visualizations are centered above the contributing physical objects, and visualizations are constrained within a display boundary to prevent unintended cropping.¹

6.1.6 Presenter Support. To reduce cognitive load and potential misuse of commands, *InSituTale* provides presenters with a private interface that displays key contextual information for each scene (see right panels in Fig. 5). This interface presents mappings between physical objects and their associated visualizations, a list of active visualization commands, and visualizations registered for transformation and composition.

6.2 Authoring Mode

InSituTale allows presenters to define scenes, specify target physical objects, assign associated visualizations, and select interaction commands for their presentations. To support iterative testing during authoring and ensure reproducibility, *InSituTale* also allows users to import templates and edit previously saved settings. The authoring interface is organized into three panels: the Interaction Setup panel (Fig. 4-B-a), the Visualization Setup panel (Fig. 4-B-b), and the Object–Visualization Mapping panel (Fig. 4-B-c). The authoring process begins with structuring the presentation into a sequence of scenes. Each scene includes a set of associated visualizations, physical objects, and visualization commands. Presenters navigate through these scenes during the live presentation, with the flexibility to move forward or backward as needed. In the Interaction Setup panel (Fig. 4-B-a), presenters enable specific visualization commands and select the preferred hand (left or right) for pointing gestures. They can also define text-based conditions to trigger visualization transformations (e.g., “Is the glass filled with red wine?”). In the Visualization Setup panel (Fig. 4-B-b), presenters select physical objects from a predefined class set and assign each object a unique identifier for runtime tracking. While *InSituTale* does not support custom visualization authoring, presenters can link physical objects to preconfigured Unity visualization prefabs (e.g., pie, bar, line, and radar charts). Annotation images and texts can also be uploaded via a web interface. All mappings and configurations are summarized in the Object–Visualization Mapping panel (Fig. 4-B-c) for easy review and management.

6.3 Implementation

We implemented *InSituTale* in Unity. The system captures the physical environment using a depth camera, analyzes the video feed in real time, and augments the physical environment with visualizations. Fig. 6 illustrates the overall system flow.

6.3.1 Object Tracking. *InSituTale* captures RGB-D video streams, enabling simultaneous object recognition and spatial tracking in 3D space. We employ a YOLOv4-based object detection model

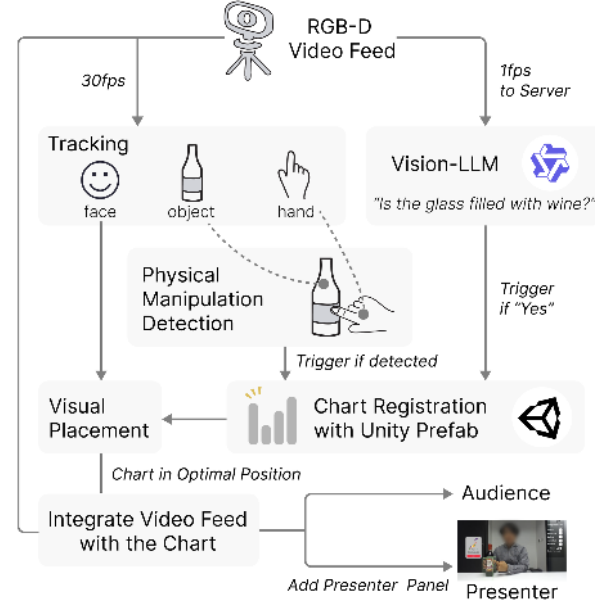


Figure 6: System Flow: *InSituTale* consists of object, hand, and face tracking, Vision-LLM, and visualization placement optimization.

trained on 80 categories from the COCO dataset [35], allowing class-level detection of physical objects. The detection model can be replaced to accommodate domain-specific object types. Detected objects are classified at the class level and assigned temporary IDs based on the order of detection within each frame. Persistent tracking across frames is achieved via a heuristic matching algorithm that associates new detections with previously tracked instances based on spatial proximity and positional continuity. To detect Compose/Decompose interactions, the system identifies the relative distances between physical objects. The system also employs plane detection to establish a baseline surface, allowing the recognition of object-lifting gestures by monitoring vertical displacement from the baseline surface. Additionally, the visualization scaling is dynamically calculated based on the distance between objects and the camera.

6.3.2 Query-Based Recognition. To enable the detection of real-world state changes aligned with narrative contexts, we incorporated a large vision-language model for real-time video analysis. We employed Qwen-VL-Chat [4], an open-source, instruction-tuned vision-language model. Each second, the system processes one compressed video frame via socket communication to this model. Frames are analyzed alongside tailored textual prompts (e.g., “Is the glass filled with red wine? Respond 1 if yes, 0 if no.”) to detect context-specific events defined during authoring.

6.3.3 Hand, Finger, and Face Tracking. Hand and finger tracking are implemented using the LightBuzz Body Tracking SDK.¹ The system distinguishes left and right hands and continuously tracks the position of the index finger on the registered side in 2D screen space. By correlating the index finger position with detected objects and visualizations, *InSituTale* recognizes pointing gestures. A face area is also detected using the Haar Cascade method [65], allowing a layout algorithm to prevent visualizations from occluding the presenter’s face.

¹<https://handtracking.lightbuzz.com>



Figure 7: Use Cases: (A) Comparing the characteristics of Japanese and Australian wines. (B) Highlighting the global market share of EV models. (C) Explaining the detailed consumption breakdown of oranges.

6.3.4 Visualization Rendering and Layout. *InSituTale* supports both 2D screen space and 3D physical space visualizations, implemented using Graph And Chart Unity Assets². To ensure visualizations and annotations remain clearly visible without occlusions, we employ a greedy algorithm informed by layout optimization methods [2, 21]. The algorithm evaluates eight candidate positions around a physical object with a tailored objective function. This function penalizes overlaps with the presenter’s face, other objects, and existing visualizations while rewarding positions aligned with the top side of the object and positions consistent with prior frames to avoid frequent and radical movements of visualizations. The candidate position achieving the highest objective score is selected. Position smoothing through linear interpolation further stabilizes visualization movements.

7 Use Cases

InSituTale supports a variety of augmented physical data storytelling scenarios (Fig. 7). We illustrate three representative examples (see the supplemental video).

7.1 Promotion Presentation

In a virtual wine-tasting session, a sommelier uses *InSituTale* to introduce and compare two wines (Fig. 7-A). Placing the first bottle on the table triggers a radar chart that visualizes key characteristics such as brand recognition, food pairing compatibility, and pricing. Pointing at the bottle helps orient the audience’s attention and reveals an annotation featuring a photo of the vineyard and a brief historical description. The second wine is placed similarly with its own radar chart and annotation. *InSituTale* automatically arranges the visuals to avoid occlusion with the presenter or other elements in the scene. To compare the wines, the sommelier brings the two bottles close together, prompting the system to overlay their radar charts and highlight similarities and differences (Fig. 5-F4). In the next scene, advanced via a ring mouse, the system shows a multi-series line chart showing evaluation scores across vintages for both wines alongside an industry benchmark (Fig. 5-D1). After explaining overall trends, the sommelier responds to a viewer’s request by improvisationally highlighting the line corresponding to the Australian wine (Fig. 5-D2), making it easier for the audience to follow. Finally, the sommelier opens one of the bottles and pours wine into a glass. This action, linked to the system prompt “*Is the glass filled with wine?*”, triggers a new radar chart displaying detailed tasting notes, including aroma, flavor, and texture (Fig. 5-G2), offering the audience a richer and more immersive tasting experience.

7.2 Product Comparison Presentation

In a virtual product showcase, a presenter compares electric vehicles (EVs), hybrid cars, and gasoline-powered vehicles using small-scale car models (Fig. 7-B). The presentation begins with a world map where dot sizes represent sales volumes for each vehicle type. The presenter first lifts the EV model, directing the audience’s attention to that category and simultaneously highlighting the corresponding dots on the map. In the next scene, two bar charts appear side by side, showing regional sales for EVs and hybrids. Moving the vehicle models horizontally causes the charts to merge into a clustered bar chart for direct comparison. Stacking the models vertically transforms the visualization into a stacked bar chart, representing the combined market share and illustrating shifts from gasoline to electric drivetrains. In the next scene, placing each model individually triggers a pie chart that breaks down cost components such as manufacturing, battery, maintenance, and fuel. Bringing the EV model closer to the camera enlarges the chart and reveals finer segments, including details like battery sourcing and warranty coverage, enabling a more nuanced cost analysis.

7.3 Consumer Behavior Presentation

In a presentation on global fruit consumption, a speaker uses an orange and a banana to illustrate market trends (Fig. 7-C). Placing the orange on the table triggers a pie chart showing its consumption breakdown—fresh use, juice production, and processed products. Similarly, placing the banana reveals its usage distribution, such as in smoothies, baby food, and fresh consumption. Pointing at each fruit brings up annotations: for example, the banana’s annotation shows an image and caption highlighting its use in school lunch programs and the logistical advantages of its protective peel, which reduces bruising and waste. Later in the presentation, the speaker peels the banana in front of the camera (Fig. 1). This action, recognized by the prompt “*Is the banana peeled?*”, updates the annotation with information about ripeness levels and glycemic index, illustrating how sensory and nutritional properties change as the fruit matures.

8 Evaluation

We conducted a user study to evaluate how presenters experience delivering data stories using *InSituTale*, focusing on its usability, utility, and learnability in real-time storytelling contexts.

8.1 Participants

We recruited 12 participants (P1–P12) with diverse backgrounds, including six males and six females, ranging in age from 19 to 30. Most had experience delivering data-centric presentations,

²<https://assetstore.unity.com/packages/tools/gui/graph-and-chart-data-visualization-78488>

though their familiarity with augmented reality varied from none to extensive. All participants attended the study in person.

8.2 Apparatus

The system ran on a laptop equipped with an Intel Core i7 processor (3.6 GHz), 32 GB RAM, and an NVIDIA RTX 3070 GPU. A separate RTX 3090 GPU hosted the vision-language model. A ZED Mini stereo camera, mounted on a tripod 50 cm in front of the presenter, captured RGB-D video input. Participants sat at a table with physical props placed in front of them. The instructor monitored the system output on a remote display. With this setup, the average latency from frame capture to visual change was 0.1 seconds ($SD = 0.02$), and the response time from the LLM server averaged 1.08 seconds ($SD = 0.067$).

8.3 Tasks and Procedure

Participants were asked to use *InSituTale* to deliver a data story in which physical objects served as key narrative props. We prepared two story sets. One set focused on **Japanese and Australian wine** and included two wine bottles and a wine glass as props. The dataset contained information such as the typical composition of wines from each region (pie chart), consumption statistics across global regions (bar chart), regional market characteristics (radar chart), tasting notes (radar chart), average ratings over the past decade (line chart), and annotated images illustrating the distinctive character of each country's wines. The second story featured a comparison between **bananas and oranges**, supported by two of each fruit as physical props. This content included market share breakdowns (donut chart), preference survey results by age group (bar chart), attribute comparisons (radar chart), export volumes (line chart), retail distribution channels (bar chart), and annotated images of the fruits' countries of origin. Participants were free to determine the structure and focus of their presentations; they were not required to use all the provided materials or follow a specific order. Instead, they were encouraged to construct a coherent narrative using a subset of the content based on their preferences.

The study was comprised of three phases: training, authoring, and presentation, lasting approximately 50–60 minutes in total. In the training phase (approximately 10 minutes), participants completed a consent form and demographic questionnaire and received a guided demonstration of *InSituTale*'s core features. During the authoring phase (15–25 minutes), each participant was assigned one of the story sets. They received the relevant physical objects and a printed handout summarizing the dataset. Using the authoring interface, participants configured scenes by assigning visualizations to physical objects and enabling interaction commands. They were allowed to ask for support from the instructor as needed, including helping them configure additional visualizations. Participants also searched for annotation images online and uploaded them through the system interface if desired. In the presentation phase (10–15 minutes), participants delivered their authored presentations using *InSituTale*. A ring mouse was used to trigger scene transitions. The instructor remained available but answered questions only upon request. After completing their presentations, participants filled out a usability questionnaire. They took part in a semi-structured interview to share qualitative feedback on their experience, the system's usability, and any perceived limitations.

8.4 Results

We analyzed the responses from the usability questionnaires and semi-structured interviews. A summary of the usability ratings is shown in Fig. 8.

Perceived Value and Benefits. Participants recognized the clear value in using *InSituTale* to enhance their current presentation practices. A majority (11 out of 12) agreed that the system improved their overall storytelling experience (Fig. 8). P1, reflecting on their experience compared to traditional slideshow-based presentations, noted: *"I liked how I could improvise with visualizations during the presentation. I ended up highlighting data points and using composite charts that I hadn't originally planned. This isn't possible in regular slides."* The ability to modify visualizations in real time, along with physical object manipulations, offered presenters a more dynamic and personalized storytelling experience. Participants also appreciated the integrated view of physical objects and visualizations. P1 shared: *"When explaining granite or basalt, professors show slides alongside physical samples. With InSituTale, they could directly link information to the objects, helping students focus on the materials."* Such integrated views can effectively support storytelling in scenarios where physical objects function as essential narrative props. Moreover, participants highlighted how the system handled visual placement. Unlike conventional video conferencing platforms like Zoom, which divide the screen between the speaker and visual aids, *InSituTale* allows visualizations, physical props, and the presenter to coexist in a single unified frame. P4 commented: *"I liked how the visualizations automatically moved to avoid covering my face or other objects. It let me stay focused on presenting rather than worrying about where visuals would appear."*

On the other hand, some participants expressed a desire for greater expressiveness. P2, who frequently gives data-driven presentations, noted: *"I expected to zoom into specific sections of bar charts, but that wasn't supported,"* highlighting the need for more fine-grained control over visualizations. P3 suggested incorporating 3D models for comparison, reflecting interest in a broader range of visual elements. P7 and P8 proposed dynamic annotations (similar to pen tools in PowerPoint) that could persistently highlight areas of interest. P2 also envisioned a higher degree of physical-digital coordination: *"It would be interesting to use a robot to move physical objects in sync with the visualizations."* Similarly, P10 imagined visual properties of physical objects adapting in real time: *"I'd like to change the appearance of a physical object, like altering a label's color, to match the visualization."*

Engagement and Interaction Experience. All participants agreed that *InSituTale* provides an engaging presentation experience (Fig. 8). As P7 remarked, *"I enjoyed the physical interactions—it felt more like a performance than just clicking through slides."* Many participants also appreciated the customizable object detection and actively experimented with their own prompts (e.g., *"Is the glass filled with wine?"*, *"Is the presenter dancing?"*, *"Is the presenter eating an orange?"*). P1 noted, *"I really liked being able to define my own triggers. It was fun when a wine-drinking gesture activated a chart. I want to explore it more."* Participants generally found the interaction techniques easy to learn. Most picked them up after a single demonstration. P8 explained, *"During my wine presentation, the system's responses to my gestures with wine bottles felt intuitive. I could remember the mappings easily because they made sense."* Furthermore, having visualizations automatically respond to physical object interactions helped maintain a

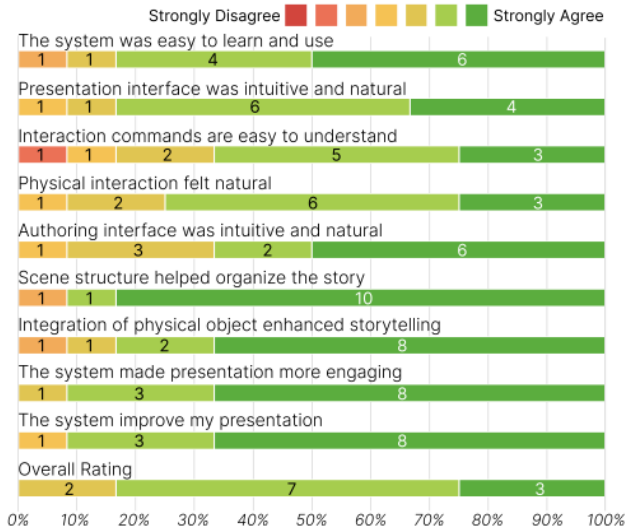


Figure 8: Usability Ratings: Participants rated the usability of *InSituTale* on a 7-point Likert scale.

smooth presentation flow. As P5 noted, “Picking up, moving closer, or pointing at objects—things I’d do anyway while explaining—automatically triggered the visuals.” Several participants also reported less physical fatigue compared to gesture-based systems. Drawing on prior VR experience, P7 remarked, “Using physical objects feels much less tiring than waving your hands in the air.”

However, a few participants noted unintended interactions. For example, P4 accidentally triggered an annotation while lifting an object due to a misinterpreted pointing gesture: “It’s hard to keep track of which hand is being used for pointing while you’re presenting.” Some objects were also harder to detect depending on how they were held. P6 noted, “The orange worked fine, but holding the banana the same way as the orange blocked detecting the banana, so I had to change my grip.” Additionally, several participants (e.g., P11) expressed interest in integrating voice commands to further extend the interaction mechanism.

Scene Management and System Behavior. Participants generally responded positively to *InSituTale*’s scene-based structure. Most participants created 4–6 scenes, typically aligning one scene per data point or topic. As P3 noted, “It felt similar to making slides—you assign each thing you want to say to its own scene.” The scene-specific presentation panel was also well-received, with participants agreeing that it supported smoother delivery. P7 shared, “Having the panel was reassuring—I didn’t need to remember which interaction triggered which visualization.”

However, some participants noted limitations in how the system handled object-to-visualization mappings—particularly when multiple objects from the same class were used. The system assigns visualizations based on the detection order, which can be unintuitive during improvisational use. P4 remarked, “The panel shows which visualization comes first, but being restricted by this sequence reduces the improvisational strength.” This limitation was particularly noticeable when object tracking failed mid-presentation, causing visualizations to mismatch. P5 and P6 described needing to remove all objects from the scene and reintroduce them in the intended order. While P5 noted, “Fixing the assignments was straightforward,” the issue nevertheless highlighted a key challenge for real-time, flexible storytelling.

Authoring Experience. Participant reactions to the authoring interface varied. P3 found the scene-based structure intuitive, noting, “Creating scenes and assigning visualizations felt similar to building slides in PowerPoint.” In contrast, P6 commented, “Configuring each scene was a bit overwhelming, especially when assigning visualizations triggered by commands (e.g., changing chart types).” Additionally, several participants framed vision-language prompts from a first-person perspective (e.g., “Am I peeling the banana?”), which reduced detection accuracy. The system performed more reliably with third-person phrasings, such as “Is the person (in the scene) peeling the banana?”, highlighting the need for clearer authoring guidance for creating prompts.

Potential Usage Scenarios. Participants proposed diverse use cases based on their backgrounds. P1 stated, “In chemistry demonstrations, it could effectively show how combining materials alters their properties.” P7 highlighted its suitability for remote sales presentations, noting, “It focuses customers’ attention on the actual products, unlike traditional slides. Sales conversations often require adapting the presentation spontaneously.” Conversely, some participants noted limitations in co-located scenarios where the presenter and audience share the same space. P5 emphasized, “Ultimately, product promotion requires letting customers physically handle the products. Therefore, visualization overlays should support co-located contexts, not just remote presentations.”

9 Discussion

We summarize design implications and potential improvements from our design process and user study.

9.1 Design Implication

Semantic Coherence Between Objects and Visuals. Aligning the semantic relationships between physical objects and their associated visualizations offers multiple benefits. Beyond enhancing visual coherence for the audience, these correspondences make interactions more intuitive and predictable—presenters can anticipate the effects of physical manipulations based on each object’s narrative role. Physical objects also serve as expressive intermediaries, affording a wider range of interactions than hand gestures alone, especially for complex or fine-grained commands. Future systems could further strengthen this coherence through bidirectional coupling between the physical and digital. For example, changes in the visualization could trigger updates in the physical environment (e.g., repositioning a physical object when a data point is selected), fostering a more fluid, engaging, and semantically rich storytelling experience.

Object-Specific Interactions. Our workshop revealed a wide variety of proposed physical manipulations, many of which were strongly tied to the affordances of specific physical objects. Supporting this diversity requires systems to account for both object-specific interpretations and more general physical manipulations that apply across a range of objects—ensuring flexibility without constraining the system to particular object types. To address this, we implemented mappings based on commonly proposed physical manipulations and integrated a vision-language model that allows presenters to define custom queries reflecting physical state changes. This feature was positively received, with participants appreciating the ability to create interactions beyond predefined gesture sets. However, we observed limitations in the system’s sensitivity to ambiguous visual changes, as well as challenges for users in formulating effective textual queries

during authoring. Future systems could improve support by providing intelligent interfaces that help presenters express intended real-world conditions in ways that are reliably detectable.

Dynamic Visualization Placement for Enhanced Visibility. Effective visualization placement is essential in dynamic presentation settings where physical elements—including the size, number, and positions of objects and presenters—continuously change. Throughout prototyping and evaluation, we observed that poor alignment often led to the occlusion of key content or disrupted audience comprehension. Our dynamic layout algorithm mitigated some of these issues by minimizing overlaps and adjusting positions in real-time. However, visual clarity can be further improved through adaptive rendering strategies—such as changing chart colors, labels, or transparency—based on object characteristics, layout constraints, or surrounding content [19, 71]. As augmented physical data storytelling moves into diverse environments (e.g., classrooms and public exhibitions), ensuring persistent visibility and coherence will be increasingly essential.

Robustness and Flexibility in Object Tracking. Our vision-based object detection pipeline provided stable tracking without the need for external electronics. However, its reliability occasionally suffered due to user handling variations (e.g., rapid or partial object movements). A key limitation arose when multiple objects of the same class were used: the heuristic ID assignment based on detection order constrained improvisational flexibility. To address this, future systems could explore hybrid tracking strategies—such as combining markerless methods with lightweight markers—to achieve more robust object differentiation while preserving usability and visual simplicity.

9.2 Limitation and Future Work

Extending Interaction Capabilities. *InSituTale* currently supports physical interactions, but future versions could benefit from enhanced multimodal capabilities to improve expressiveness and adaptability. Prior work has shown that voice and gesture-based inputs can enrich real-time presentations [22, 32]. Both workshop participants and study users suggested combining modalities, such as lifting an object while speaking, to enable more natural and flexible interaction patterns. Incorporating multimodal input would allow presenters to coordinate physical, verbal, and gestural cues in more expressive and intuitive ways.

In addition to expanding presenter input, integrating lightweight audience feedback mechanisms could further enhance communication. Monitoring real-time audience reactions remains a challenge, often limiting presenters' ability to respond effectively [40]. Adding simple feedback channels, such as reaction icons, could help presenters better assess engagement and adjust their storytelling in the moment.

Audience-Oriented Evaluation. While *InSituTale* targets real-time data presentations, our evaluation focused mainly on presenter experience and system functionality. We did not formally assess how audiences perceive or benefit from *InSituTale* compared to conventional slide-based presentations, which we see as the most relevant baseline given its focus on real-time delivery. Although some presenters noted improved clarity and engagement, systematic audience-side evaluation, such as measuring comprehension or attention across different contexts, remains important future work.

Limited Object Diversity. Our interaction design was grounded in six representative physical objects with varied affordances. While informative, this range is limited, and the identified interactions may not generalize to objects with different forms or materials. Moreover, the semantic mappings are still influenced by the chosen set of objects. A broader exploration of physical objects could uncover a wider range of interaction possibilities and improve adaptability across diverse storytelling contexts.

Focus on Remote Presentations. *InSituTale* was primarily developed for remote settings. However, several participants noted its potential in co-located environments. Future work could explore adapting the system for in-person use with projection-based AR or HMDs, similar to existing study approaches [9, 60]. This enables audiences to experience and interact with visualizations and physical objects directly.

Authoring Support. The current authoring supports of *InSituTale* are limited to using templates and previously saved settings. Participants also noted challenges in associating physical objects with appropriate visuals for each scene and narrative prompt. To enhance usability, future systems could offer intelligent recommendations and allow users to define outputs through direct manipulation of physical objects [31, 38], supporting a more intuitive and flexible authoring experience.

10 Conclusion

This study explored augmented physical data storytelling, an approach that enables presenters to control visualizations through intuitive manipulations of physical objects, seamlessly blending the physical and digital. We conducted workshops with nine VIS/HCI researchers to investigate how different types of physical actions can be mapped to visualization behaviors in narrative contexts. These insights informed the design of *InSituTale*, a prototype system that integrates physical-space object tracking via a depth camera, a vision-language model for customized detection, and dynamic visualization placement. Our evaluation with 12 participants showed that *InSituTale* enables intuitive, engaging, and expressive data storytelling through cohesive integration of physical and digital elements. We hope this work encourages further exploration of physical-object-based interaction in data-driven presentation systems.

Acknowledgments

We thank the reviewers for their valuable feedback. We are also especially grateful to Leni Yang at INRIA Bordeaux for the insightful discussions. This work was supported by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF) grant 16214623, the Knut and Alice Wallenberg Foundation under Grant KAW 2019.0024, and JST PRESTO Grant Number JP-MJPR23I5. Takanori Fujiwara completed this work while affiliated with Linköping University.

References

- [1] The Visual Agency. 2014. The Visual Agency - Annual Report 2014. <https://vimeo.com/123407907>.
- [2] R. Azuma and C. Furmanski. 2003. Evaluating label placement for augmented reality view management. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings*. IEEE Computer Society, Los Alamitos, CA, USA, 66–75. doi:10.1109/ISMAR.2003.1240689
- [3] S. Sandra Bae, Takanori Fujiwara, Anders Ynnerman, Ellen Yi-Luen Do, Michael L. Rivera, and Danielle Albers Szafr. 2024. A Computational Design Pipeline to Fabricate Sensing Network Physicalizations. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 913–923. doi:10.1109/TVCG.2023.3327198

- [4] Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond. arXiv:2308.12966 [cs.CV] <https://arxiv.org/abs/2308.12966>
- [5] Matthew Brehmer and Robert Kosara. 2022. From Jam Session to Recital: Synchronous Communication and Collaboration Around Data in Organizations. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 1139–1149. doi:10.1109/TVCG.2021.3114760
- [6] Yining Cao, Rubaiat Habib Kazi, Li-Yi Wei, Deepali Aneja, and Haijun Xia. 2024. Elastica: Adaptive Live Augmented Presentations with Elastic Mappings Across Modalities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 599, 19 pages. doi:10.1145/3613904.3642725
- [7] Zhutian Chen, Daniele Chiappalupi, Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2023. RL-L: A Deep Reinforcement Learning Approach Intended for AR Label Placement in Dynamic Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (Oct. 2023), 1347–1357. doi:10.1109/TVCG.2023.3326568
- [8] Paul Crowther. 2019. *Theory of the Art Object*. Routledge, New York.
- [9] Peter Dalsgaard and Kim Halskov. 2012. Tangible 3D tabletops: combining tangible tabletop interaction and 3D projection. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) (NordiCHI '12). Association for Computing Machinery, New York, NY, USA, 109–118. doi:10.1145/2399016.2399033
- [10] Josh Urban Davis, Paul Asente, and Xing-Dong Yang. 2023. Multimodal Direct Manipulation in Video Conferencing: Challenges and Opportunities. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 1174–1193. doi:10.1145/3563657.3596099
- [11] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Coiaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST '24). Association for Computing Machinery, New York, NY, USA, Article 19, 15 pages. doi:10.1145/3654777.3676379
- [12] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. 2015. Situated Analytics. In *2015 Big Data Visual Analytics (BDVA)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–8. doi:10.1109/BDVA.2015.7314278
- [13] Neven A.M. ElSayed, Bruce H. Thomas, Kim Marriott, Julia Piantadosi, and Ross T. Smith. 2016. Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages & Computing* 36 (2016), 13–23. doi:10.1016/j.jvlc.2016.07.006
- [14] Barrett Ens, Sarah Goodwin, Arnaud Prouzeau, Fraser Anderson, Florence Y. Wang, Samuel Gratzl, Zac Lucarelli, Brendan Moyle, Jim Smiley, and Tim Dwyer. 2021. Uplift: A Tangible and Immersive Tabletop System for Casual Collaborative Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1193–1203. doi:10.1109/TVCG.2020.3030334
- [15] Temiloluwa Paul Femi-Gege, Matthew Brehmer, and Jian Zhao. 2024. VisConductor: Affect-Varying Widgets for Animated Data Storytelling in Gesture-Aware Augmented Video Presentation. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 531 (Oct. 2024), 22 pages. doi:10.1145/3698131
- [16] Alba Fombaro, Guillermo Fombaro, and Octavio Fombaro. 2022. Augmented Reality, a Review of a Way to Represent and Manipulate 3D Chemical Structures. *Journal of Chemical Information and Modeling* 62, 8 (2022), 2012–2028. doi:10.1021/acs.jcim.1c01255
- [17] Gapminder Foundation. 2014. The Future of Ebola if Not Stopped Now - Rosling's Factpod #8. https://www.youtube.com/watch?v=sYVUs2_F5Kg.
- [18] Mikhaila Friske, Jordan Wirfs-Brock, and Laura Devendorf. 2020. Entangling the Roles of Maker and Interpreter in Interpersonal Data Narratives: Explorations in Yarn and Sound. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) (DIS '20). Association for Computing Machinery, New York, NY, USA, 297–310. doi:10.1145/3357236.3395442
- [19] Zeinab Ghaemi, Kadek Ananta Satriadi, Ulrich Engelke, Barrett Ens, and Bernhard Jenny. 2023. Visualization Placement for Outdoor Augmented Data Tours. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) (SUI '23). Association for Computing Machinery, New York, NY, USA, Article 9, 14 pages. doi:10.1145/3607822.3614518
- [20] Weilun Gong, Stephanie Santosa, Tovi Grossman, Michael Glueck, Daniel Clarke, and Frances Lai. 2023. Affordance-Based and User-Defined Gestures for Spatial Tangible Interaction. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 1500–1514. doi:10.1145/3563657.3596032
- [21] Raphaël Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. 2012. Image-driven view management for augmented reality browsers. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Los Alamitos, CA, USA, 177–186. doi:10.1109/ISMAR.2012.6402555
- [22] Brian D. Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented Chironomia for Presenting Data to Remote Audiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22). Association for Computing Machinery, New York, NY, USA, Article 18, 14 pages. doi:10.1145/3526113.3545614
- [23] Shuqi He, Haonan Yao, Luyan Jiang, Kaiwen Li, Nan Xiang, Yue Li, Haining Liang, and Lingyun Yu. 2024. Data Cubes in Hand: A Design Space of Tangible Cubes for Visualizing 3D Spatio-Temporal Data in Mixed Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 209, 21 pages. doi:10.1145/3613904.3642740
- [24] Bridger Herman, Maxwell Omdal, Stephanie Zeller, Clara A. Richter, Francesca Samsel, Greg Abram, and Daniel F. Keefe. 2021. Multi-Touch Querying on Data Physicalizations in Immersive AR. *Proc. ACM Hum.-Comput. Interact.* 5, ISS, Article 497 (Nov. 2021), 20 pages. doi:10.1145/3488542
- [25] Mark Hough. 2017. Adidas - Low Waste Brand Film. <https://vimeo.com/199539760>.
- [26] Flow Immersive. 2024. Life Expectancy Avatar Story. <https://www.youtube.com/watch?v=MMB3rHeZdcE>.
- [27] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI '97). Association for Computing Machinery, New York, NY, USA, 234–241. doi:10.1145/258549.258715
- [28] Terry Haekyung Kim and Ho Jung Choo. 2021. Augmented reality as a product presentation tool: focusing on the role of product information and presence in AR. *Fashion and Textiles* 8, 1 (2021), 1–21. doi:10.1186/s40691-021-00261-w
- [29] Adrian Kristanto, Maxime Cordeil, Benjamin Tag, Nathalie Henry Riche, and Tim Dwyer. 2023. Hanstreamer: an Open-source Webcam-based Live Data Presentation System. arXiv:2309.12538 [cs.HC] <https://arxiv.org/abs/2309.12538>
- [30] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. 2013. SketchStory: Telling More Engaging Stories with Data through Freeform Sketching. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2416–2425. doi:10.1109/TVCG.2013.191
- [31] Boyu Li, Linping Yuan, Zhe Yan, Qianxi Liu, Yulin Shen, and Zeyu Wang. 2024. AniCraft: Crafting Everyday Objects as Physical Proxies for Prototyping 3D Character Animation in Mixed Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST '24). Association for Computing Machinery, New York, NY, USA, Article 99, 14 pages. doi:10.1145/3654777.3676325
- [32] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-Time Speech-Driven Augmented Presentation for AR Live Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22). Association for Computing Machinery, New York, NY, USA, Article 17, 12 pages. doi:10.1145/3526113.3545702
- [33] Jian Liao, Kevin Van, Zhijie Xia, and Ryo Suzuki. 2024. RealityEffects: Augmenting 3D Volumetric Videos with Object-Centric Annotation and Dynamic Visual Effects. In *Proceedings of the 35th Annual ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 1248–1261. doi:10.1145/3643834.3661631
- [34] Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2023. Labeling Out-of-View Objects in Immersive Analytics to Support Situated Visual Searching. *IEEE Transactions on Visualization and Computer Graphics* 29, 3 (2023), 1831–1844. doi:10.1109/TVCG.2021.3133511
- [35] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
- [36] Jingyuan Liu, Hongbo Fu, and Chiew-Lan Tai. 2020. PoseTween: Pose-driven Tween Animation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '20). Association for Computing Machinery, New York, NY, USA, 791–804. doi:10.1145/3379337.3415822
- [37] Weizhou Luo, Zhongyuan Yu, Rufat Rzayev, Marc Satkowski, Stefan Gumhold, Matthew McGinity, and Raimund Dachselt. 2023. Pearl: Physical Environment based Augmented Reality Lenses for In-Situ Human Movement Analysis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 381, 15 pages. doi:10.1145/3544548.3580715
- [38] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 459, 15 pages. doi:10.1145/3544548.3581449
- [39] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *Interactions* 21, 3 (may 2014), 40–45. doi:10.1145/2591689
- [40] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. AffectiveSpotlight: Facilitating the Communication of Affective Responses from Audience Members during Online Presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 247, 13 pages. doi:10.1145/3411764.3445235

- [41] Biswaksen Patnaik, Huaishu Peng, and Niklas Elmqvist. 2024. VisTorch: Interacting with Situated Visualizations using Handheld Projectors. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 208, 13 pages. doi:10.1145/3613904.3642857
- [42] Ken Perlin, Zhenyi He, and Karl Rosenberg. 2018. Chalktalk : A Visualization and Communication Language – As a Tool in the Domain of Computer Science Education. arXiv:1809.07166 [cs.HC] <https://arxiv.org/abs/1809.07166>
- [43] Dominic Potts, Martynas Dabravalskis, and Steven Houben. 2022. Tangible-Touch: A Toolkit for Designing Surface-based Gestures for Tangible Interfaces. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '22). Association for Computing Machinery, New York, NY, USA, Article 39, 14 pages. doi:10.1145/3490149.3502263
- [44] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. 2018. *Data-Driven Storytelling*. CRC Press, Boca Raton, FL. doi:10.1201/9781315281575
- [45] Dario Rodighiero, Lins Derry, Douglas Duhaime, Jordan Kruguer, Maximilian C Mueller, Christopher Pietsch, Jeffrey T Schnapp, and Jeff Steward. 2022. Surprise machines: Revealing harvard art museums' image collection. *Information Design Journal* 27, 1 (2022), 21–34.
- [46] Hans Rosling. 2007. The best stats you've ever seen. <https://www.youtube.com/watch?v=hVimVzgtD6w>.
- [47] Hans Rosling. 2013. DON'T PANIC — Hans Rosling showing the facts about population. <https://vimeo.com/79878808>.
- [48] Hans Rosling. 2013. The River of Myths. <https://www.youtube.com/watch?v=IypX4l2UeZg>.
- [49] Hans Rosling. 2014. Global population growth, box by box. https://www.ted.com/talks/hans_rosling_global_population_growth_box_by_box.
- [50] Hans Rosling. 2016. Hans Rosling: What if every squash had a washing machine? https://www.youtube.com/watch?v=rMPCQ0XI25c&ab_channel=BillGates.
- [51] Hans Rosling. 2016. Numbers are boring, people are interesting. https://www.youtube.com/watch?v=nh94kK05l-M&ab_channel=TEDxTalks.
- [52] Hans Rosling. 2016. Why the world population won't exceed 11 billion. <https://www.youtube.com/watch?v=2LyzBoHo5EI>.
- [53] Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, and Wilmot Li. 2019. Interactive Body-Driven Graphics for Augmented Video Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300852
- [54] Kadek Ananta Satriadi, Barrett Ens, Sarah Goodwin, and Tim Dwyer. 2023. Active Proxy Dashboard: Binding Physical Referents and Abstract Data Representations in Situated Visualization through Tangible Interaction. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 23, 7 pages. doi:10.1145/3544549.3585797
- [55] Kadek Ananta Satriadi, Jim Smiley, Barrett Ens, Maxime Cordeil, Tobias Czuderna, Benjamin Lee, Ying Yang, Tim Dwyer, and Bernhard Jenny. 2022. Tangible Globes for Data Visualisation in Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 505, 16 pages. doi:10.1145/3491102.3517715
- [56] Orit Shaer and Eva Hornecker. 2010. Tangible User Interfaces: Past, Present, and Future Directions. *Found. Trends Hum.-Comput. Interact.* 3, 1–2 (jan 2010), 1–137. doi:10.1561/11000000026
- [57] Arjun Srinivasan and Matthew Brehmer. 2023. Combining Voice and Gesture for Presenting Data to Remote Audiences. In *IEEE VIS 2023 Workshop on Multimodal Experiences for Remote Communication Around Data Online, MERCADO'23*. IEEE Computer Society, Los Alamitos, CA, USA.
- [58] Critical Statistics. 2019. Better and better? A comment on Hans Rosling. https://www.youtube.com/watch?v=OoLcsj9ysvs&ab_channel=CriticalStatistics.
- [59] Adam Strantz. 2023. Print-and-play: Data Physicalization Methods for Research Analysis in Technical Communication. In *Proceedings of the 41st ACM International Conference on Design of Communication* (Orlando, FL, USA) (SIG-DOC '23). Association for Computing Machinery, New York, NY, USA, 215–220. doi:10.1145/3615335.3623039
- [60] Kento Takahira, Wong Kam-Kwai, Leni Yang, Xian Xu, Takanori Fujiwara, and Huamin Qu. 2025. TangibleNet: Synchronous Network Data Storytelling through Tangible Interactions in Augmented Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 233, 18 pages. doi:10.1145/3706598.3714265
- [61] Pratik Tarafdar, Alvin Chung Man Leung, Wei Thoo Yue, and Indranil Bose. 2024. Understanding the impact of augmented reality product presentation on diagnosticity, cognitive load, and product sales. *International Journal of Information Management* 75 (2024), 102744. doi:10.1016/j.ijinfomgt.2023.102744
- [62] Wai Tong, Chen Zhu-Tian, Meng Xia, Leo Yu-Ho Lo, Linping Yuan, Benjamin Bach, and Huamin Qu. 2023. Exploring Interactions with Printed Data Visualizations in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 418–428. doi:10.1109/TVCG.2022.3209386
- [63] B. Ullmer and H. Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM Systems Journal* 39, 3.4 (2000), 915–931. doi:10.1147/sj.393.0915
- [64] John Underkoffler and Hiroshi Ishii. 1999. Urp: a luminous-tangible workbench for urban planning and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) (CHI '99). Association for Computing Machinery, New York, NY, USA, 386–393. doi:10.1145/302979.303114
- [65] Paul Viola and Michael J Jones. 2004. Robust real-time face detection. *International journal of computer vision* 57 (2004), 137–154. doi:10.1023/B:VISI.0000013087.49260.fb
- [66] Vox. 2015. Obama on what most Americans get wrong about foreign aid. https://www.youtube.com/watch?v=nzL_avUIIEE.
- [67] Zhen Wen, Wei Zeng, Luoxuan Weng, Yihan Liu, Mingliang Xu, and Wei Chen. 2023. Effects of View Layout on Situated Analytics for Multiple-View Representations in Immersive Visualization. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 440–450. doi:10.1109/TVCG.2022.3209475
- [68] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. 2017. Embedded Data Representations. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 461–470. doi:10.1109/TVCG.2016.2598608
- [69] Lijie Yao, Federica Bucchieri, Victoria McArthur, Anastasia Bezerianos, and Petra Isenberg. 2025. User Experience of Visualizations in Motion: A Case Study and Design Considerations. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 174–184. doi:10.1109/TVCG.2024.3456319
- [70] Xiaoyan Zhou, Yalong Yang, Francisco Ortega, Anil Ufuk Batmaz, and Benjamin Lee. 2023. Data-driven Storytelling in Hybrid Immersive Display Environments. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE Computer Society, Los Alamitos, CA, USA, 242–246. doi:10.1109/ISMAR-Adjunct60411.2023.00056
- [71] Chen Zhu-Tian, Daniele Chiappalupi, Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2024. RL-L: A Deep Reinforcement Learning Approach Intended for AR Label Placement in Dynamic Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (Jan. 2024), 1347–1357. doi:10.1109/TVCG.2023.3326568