DataMagic: Transforming Tabular Data into Data Insight Video

Yupeng Xie
HKUST (GZ)
yxie740@connect.hkust-gz.edu.cn

Chen Ma
China Unicom
machen2@chinaunicom.cn

Zhenyang Wang
HKUST (GZ)
wwwangzhenyang@gmail.com

Liangwei Wang
HKUST (GZ)
lwang344@connect.hkust-gz.edu.cn

Jiayi Zhu
HKUST (GZ)
jzhu351@connect.hkust-gz.edu.cn

Chuxuan Zeng
China Unicom
zengchuxuan@chinaunicom.cn

Zhouan Shen
HKUST (GZ)
zshen575@connect.hkust-gz.edu.cn

Boyan Li
HKUST (GZ)
bli303@connect.hkust-gz.edu.cn

Yuyu Luo*
HKUST (GZ)
yuyuluo@hkust-gz.edu.cn

ABSTRACT

Data videos integrate dynamic charts, voice narration, and synchronized animations to communicate data insights as temporal narratives, making them an effective medium for improving data consumption efficiency in the data management lifecycle. However, producing high-quality data videos requires expertise spanning data analysis, narrative design, and video production. Existing approaches fall short: static visualization tools (e.g., BI dashboards) lack narrative logic and animation; authoring tools require users to pre-prepare visualizations rather than working from raw data; pixel-level video generation models cannot guarantee data fidelity or provenance. We demonstrate DataMagic, an end-to-end interactive system that transforms raw tabular data and natural language queries into narrative data-insight videos. To ensure data fidelity, DataMagic introduces the declarative specification DVSpec, which binds visual and animation elements to underlying data fields through data-driven semantic references. To address the combinatorial explosion of the design space, DataMagic adopts a Generate-then-Orchestrate multi-agent architecture that generates candidate scenes in parallel and then optimizes narrative coherence through global orchestration. Leveraging DVSpec's decoupling of logic and rendering, the system further supports three interaction modes and structured provenance-based data Q&A, transforming one-way videos into explorable interactive data interfaces. Evaluation on 109 real-world samples validates the effectiveness of the DataMagic.

PVLDB Reference Format:
Yupeng Xie, Chen Ma, Zhenyang Wang, Liangwei Wang, Jiayi Zhu, Chuxuan Zeng, Zhouan Shen, Boyan Li, and Yuyu Luo. DataMagic: Transforming Tabular Data into Data Insight Video. PVLDB, 19(12): 4526 - 4529, 2026.
doi:10.14778/3827998.3828057

PVLDB Artifact Availability:
The artifacts are available at https://github.com/HKUSTDial/DataMagic.

*Yuyu Luo is the corresponding author.

Proceedings of the VLDB Endowment, Vol. 19, No. 12 ISSN 2150-8097.
doi:10.14778/3827998.3828057

1 INTRODUCTION

Data videos integrate dynamic charts, voice narration, and synchronized animations to communicate data insights as temporal narratives, attracting growing attention from both academia and industry [1, 8]. However, producing high-quality data videos requires cross-domain expertise spanning data analysis, narrative design, and video production [8], and existing approaches fall short of end-to-end automation from raw data to data video [10, 11].

Static visualization tools (e.g., BI dashboards, HAIChart [12]) output static charts, lacking narrative logic and animation. Authoring tools (e.g., Data Playwright [7]) focus on adding animation effects to existing charts, without extracting insights from raw data or handling multi-scene narrative orchestration. Pixel-level video generation models can synthesize videos, but their black-box nature frequently produces numerical hallucinations and cannot map visual elements back to underlying data records.

Our key observation is that effective data videos are fundamentally structured narratives rather than simple assemblies of visual elements [8]. This motivates us to model end-to-end generation as a hierarchical content orchestration problem. Two core challenges emerge: (1) how to design a structured intermediate representation that precisely describes heterogeneous components and their temporal relations while ensuring data fidelity and provenance; (2) how to efficiently search the vast design space for solutions that balance local scene quality and global narrative coherence.

To this end, we demonstrate DataMagic, an end-to-end interactive system from raw tabular data to narrative data videos. For challenge (1), we introduce DVSpec (Data Video Specification), a declarative specification that binds visual and animation elements to underlying data fields through data-driven semantic references and narration-index triggering, ensuring full provenance. For challenge (2), DataMagic adopts a Generate-then-Orchestrate multi-agent architecture that generates diverse candidate scenes in parallel and then applies global orchestration to optimize scene selection, ordering, and narrative coherence. Leveraging DVSpec's decoupling of logic and rendering, the system further supports three interaction modes and structured provenance-based data Q&A, enabling users to refine videos and explore data directly.

The main contributions are: (1) DataMagic, which integrates DVSpec and Generate-then-Orchestrate to ensure data fidelity and

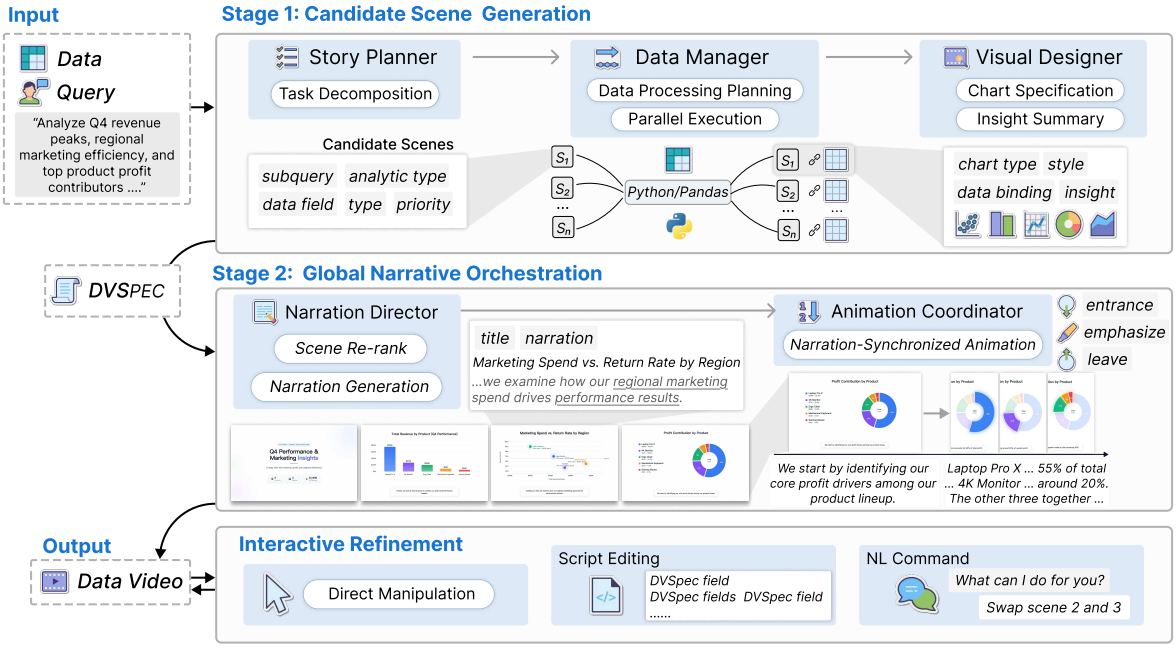

Figure 1: System architecture of DataMagic.

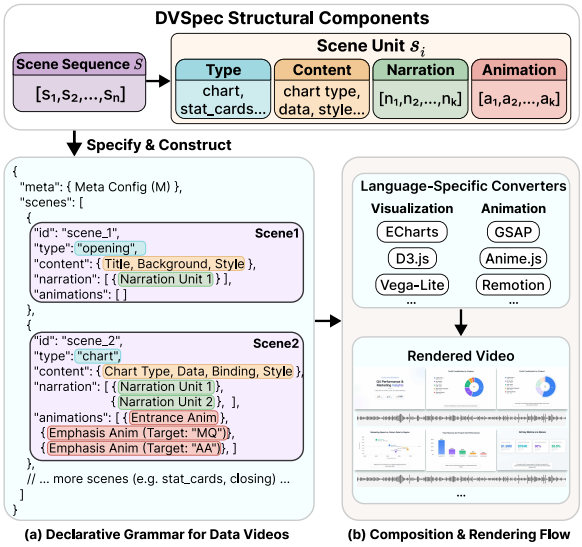

Figure 2: DVSpec structure and rendering flow.

provenance; and (2) three demonstration scenarios covering automated generation, multimodal editing, and provenance-based data Q&A, with validation on 109 real-world samples.

2 SYSTEM ARCHITECTURE

As shown in Figure 1, DataMagic transforms raw tabular datasets and natural language queries into narrative videos through a multi-agent engine, a DVSpec intermediate representation, and a rendering engine. Its design consists of three components: DVSpec, multi-agent generation, and provenance-based interaction.

2.1 DVSpec: Declarative Data Video Specification

Existing declarative specifications [4, 6] perform well for static charts or single-chart annotations and animations, but have not been extended to the unified description of cross-modal content and temporal coordination required for multi-scene data videos [2, 3]. To fill this gap, we design DVSpec (Data Video Specification), a rendering-library-agnostic declarative specification that serves as a structured intermediate representation between the multi-agent generation engine and the rendering engine. Generation results are written to DVSpec, while interactive edits are mapped to local DVSpec updates. Our implementation renders charts with D3.js and synthesizes videos with Remotion.

As shown in Figure 2(a), DVSpec formalizes a data video as a combination of metadata M and an ordered scene sequence $S = \langle s_1, \ldots, s_n \rangle$: $V := (M, S)$. Each scene s_i is defined as a four-tuple $s_i := (type, content, narration, animation)$, where *content* encapsulates visualization configuration (chart type, data bindings, style parameters), *narration* is an ordered list of narration segments, and *animation* is a list of animation effects. The scene-based design stems from research on data video narrative structures [1, 9].

DVSpec introduces two key mechanisms to ensure data fidelity and audio-video synchronization:

(1) Data-driven semantic references. Visual elements are referenced by data attribute values (e.g., `{"company": "Nvidia"}`) rather than hard-coded identifiers, keeping references stable across data updates or chart-type changes and traceable to tabular records. Fuzzy matching handles minor naming differences at render time, while retries handle code-execution failures.

(2) Narration-index declarative triggering. Animation trigger timing is declared using narration segment indices rather than absolute timestamps. During rendering, the system automatically aligns animations based on the actual audio duration generated by text-to-speech (TTS), so that when users modify narration text, animation synchronization relationships are automatically maintained without manual keyframe adjustment [9].

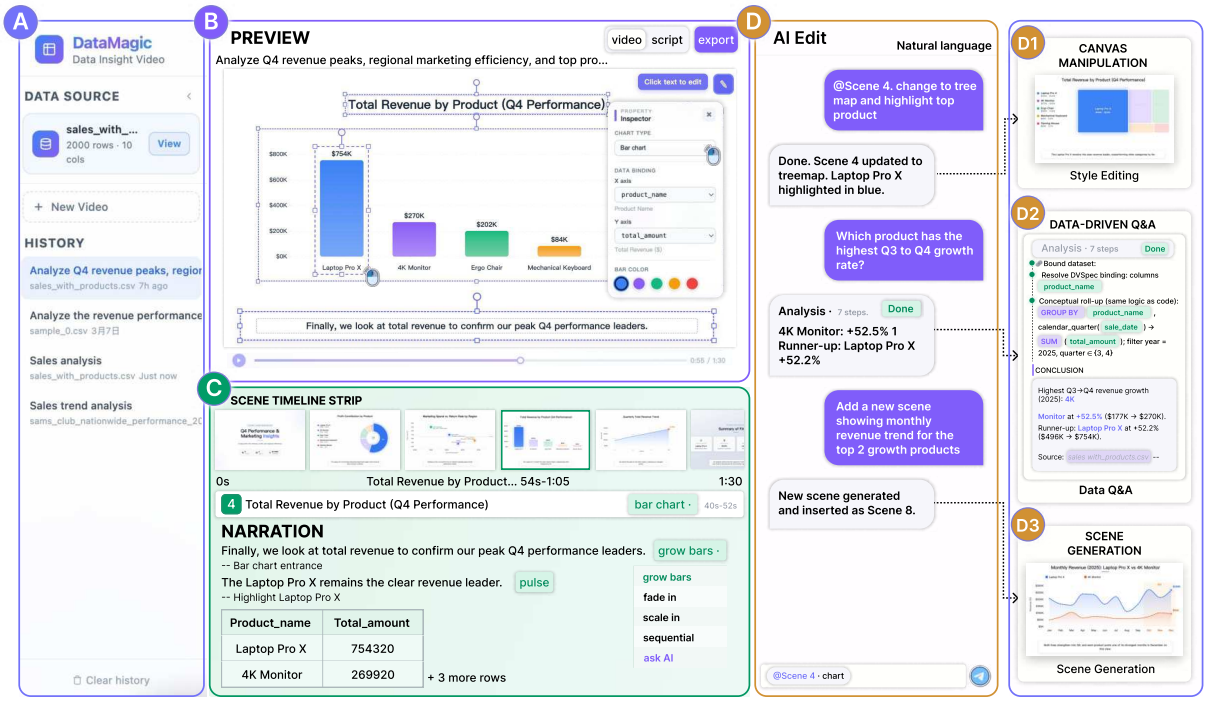

Figure 3: DataMagic **web interface with the main UI components (A-D), (D1-D3) highlight key interaction flows.**

2.2 Multi-Agent Generation Pipeline

Data video generation tightly couples data processing, visual design, and narrative logic, making it difficult for single-stage methods to ensure both per-scene accuracy and global coherence. DataMagic therefore adopts a Generate-then-Orchestrate architecture: it generates candidate scenes in parallel and then globally optimizes their selection, ordering, and transitions.

Stage 1: Data-driven candidate generation. The Story Planner decomposes the user's high-level query into several independent analytical sub-tasks based on the analytical dimensions involved (e.g., temporal trends, category comparisons, regional distributions). For each sub-task, the Data Manager plans the data processing workflow and automatically generates Python code to extract, filter, and aggregate relevant data slices from the source table. The Visual Designer then designs the visualization scheme (chart type, data bindings, and style configuration), extracts key insights, and populates the content field of DVSpec. This stage generates a pool of candidate scenes in parallel, decoupling low-level data processing from high-level narrative construction.

Stage 2: Global narrative orchestration. The Narration Director selects scenes from the candidate pool based on insight value and query coverage, and plans a playback order following narrative logic (e.g., "macro-to-micro", "phenomenon before cause"). It generates narration for each scene conditioned on adjacent scenes' context, ensuring narrative coherence across scene transitions. The Animation Coordinator uses DVSpec's semantic reference mechanism to bind data entities mentioned in the narration to corresponding visual elements, achieving end-to-end audio-video synchronization.

The generated DVSpec is compiled into the final video by the rendering engine, with narration text synthesized into speech via TTS. The system adopts a model-agnostic design to support different LLM backends.

2.3 Provenance-Based Interaction and Exploration

Prior work has established data provenance as a key enabler for interactive visualizations [5]. DataMagic extends this idea to the data video setting: DVSpec records the binding relationships from visual elements to underlying data fields during the generation stage. This structured data provenance information permeates the system's downstream interactions, supporting two core capabilities.

Parametric Editing. By linking data fields and narration segments to visual and animation elements, DVSpec localizes user edits without regenerating the full video. Users can refine a video through canvas manipulation (e.g., switching chart types or adjusting data mappings), script editing (e.g., modifying narration or animation parameters), and natural language commands (e.g., "@Scene 4: change to treemap and highlight top product"). All modes share one DVSpec state for consistency.

Data Q&A. The provenance relationships preserved by DVSpec enable the system to map users' natural language questions to specific data binding fields: the system uses an LLM to parse the question, identifies relevant data fields from the current scene's DVSpec bindings, and constructs structured query operations (e.g., filtering, aggregation, extremum retrieval) to execute directly on the underlying tabular data, rather than relying on visual models to infer content from pixels. Users can also transform insights discovered through Q&A into new video scenes, completing the full loop from data exploration to narrative expansion.

3 DEMONSTRATION SCENARIOS

We design three demonstration scenarios to showcase DataMagic's capabilities in end-to-end data video generation, DVSpec-based multimodal editing, and provenance-based data Q&A. As shown in Figure 3, the web interface integrates data management, video preview, scene timeline, narration editing, and AI-assisted interaction, allowing users to inspect and refine generated data videos within a unified workflow.

3.1 Data-Driven Automated Generation

This scenario demonstrates how DataMagic automatically transforms raw tabular data and natural language queries into narrative data videos. After a user uploads a business CSV file and specifies an analysis goal, the system performs query decomposition, data processing, chart design, and global narrative orchestration. The generated video is played in the preview panel, while the corresponding scene timeline, narration, animation labels, and related data tables are shown to help users inspect the bindings between each video segment and the underlying data. The history panel records intermediate outputs for traceability.

3.2 DVSpec-Based Multimodal Editing

This scenario demonstrates iterative post-generation editing of the generated data video. Since DVSpec stores visual elements, narration, and animations in a structured form, users can interleave canvas operations, script edits, and natural language commands on the same video state. For example, a user may convert a bar-chart scene into a treemap and highlight the top product; DataMagic then updates the corresponding DVSpec fields and incrementally re-renders the scene. Narration changes are automatically realigned with animations through the narration-index mechanism. This demonstrates specification-level editing rather than pixel-level video manipulation.

3.3 Structured Provenance-Based Data Q&A

This scenario demonstrates how DataMagic turns data videos into queryable and extensible interactive data interfaces. The data bindings and provenance relationships preserved in DVSpec allow the system to translate user questions about the current scene or the entire video into structured data queries, rather than inferring answers from video pixels. This enables multi-granularity exploration from individual scenes to the full video. For instance, when a user asks which product has the highest Q3-to-Q4 growth rate, the system resolves the relevant fields, queries the original dataset, and returns concrete results such as "4K Monitor: +52.5%" and the runner-up "Laptop Pro X: +52.2%". Users can further ask DataMagic to add a scene showing the monthly revenue trend of the top-growth products, completing the loop from data exploration to narrative expansion.

3.4 Quantitative Evaluation

We evaluated DataMagic on 109 real-world samples from DACompDA and T2R-bench. We first used four LLMs (DeepSeek-V3.2, Gemini-2.5-Pro, GPT-5, and Claude-Sonnet-4) to directly generate data videos in one pass, probing current capability limits; we also integrated these models into the DataMagic framework to verify model-agnosticism. The evaluation covered execution rate (the fraction of samples for which the full pipeline completes without error and renders a playable video) and five quality dimensions (Intent Fulfillment, Data Insights, Narrative Quality, Animation Effectiveness, Aesthetic Quality; 1–5 scale) [13]. Gemini-2.5-Pro served as the judge (r=0.91 vs. human experts on 60 videos).

Results show that even strong LLMs struggle to coordinate data analysis, narration, and animation in one-pass generation: quality scores ranged from 1.91 to 2.22, with execution rates between 48–86%. Common failures were audio-visual misalignment and weak narrative structure, where narration fell out of sync with visuals and scenes lacked logical transitions. DataMagic addressed these problems through DVSpec's narration-index triggering mechanism and the Generate-then-Orchestrate strategy, raising the average score to 3.89 with execution rates above 95%. The largest gains were in Animation Effectiveness (1.84→4.03, +120%) and Narrative Quality (2.00→3.43, +72%). Direct generation averaged ~57 seconds, while DataMagic took ~176 seconds (60 s configuration generation + 116 s rendering at parallelism = 3), reflecting a quality–latency trade-off. Ablations confirm both components' importance: removing Story Planner lowers quality by 11.6% (Narrative −15.1%), and removing Orchestration by 9.0% (Intent −12.4%).

ACKNOWLEDGMENTS

This paper was supported by the NSF of China (62402409); Youth S&T Talent Support Programme of Guangdong Provincial Association for Science and Technology (SKXRC2025461); the Young Talent Support Project of Guangzhou Association for Science and Technology (QT-2025-001); Guangzhou Basic and Applied Basic Research Foundation (2026A1515010269, 2025A04J3935, 2023A1515110545); Guangzhou-HKUST(GZ) Joint Funding Program (2025A03J3714); and Research and Development of Heterogeneous Computing Interconnection and Scheduling Software Stack (YF202400000003).

REFERENCES

[1] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, et al. 2015. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. In *CHI*. 1459–1468.
[2] Yiru Chen, Xupeng Li, Jeffrey Tao, et al. 2025. Physical Visualization Design: Decoupling Interface and System Design. *PACMMOD* 3, 3 (2025), 197:1–197:27.
[3] Yiru Chen, Jeffrey Tao, and Eugene Wu. 2023. DIG: The Data Interface Grammar. In *HILDA*. 7:1–7:7.
[4] Yiyu Chen, Yifan Wu, Shuyu Shen, et al. 2025. ChartMark: A Structured Grammar for Chart Annotation. In *IEEE VIS*. 311–315.
[5] Fotis Psallidas and Eugene Wu. 2018. Provenance for Interactive Visualizations. In *HILDA*. 9:1–9:8.
[6] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, et al. 2017. Vega-Lite: A Grammar of Interactive Graphics. *TVCG* 23, 1 (2017), 341–350.
[7] Leixian Shen, Haotian Li, Yun Wang, et al. 2025. Data Playwright: Authoring Data Videos With Annotated Narration. *TVCG* 31, 9 (2025), 5884–5897.
[8] Leixian Shen, Haotian Li, Yun Wang, et al. 2025. Reflecting on Design Paradigms of Animated Data Video Tools. In *CHI*. 190:1–190:21.
[9] Leixian Shen, Yizhi Zhang, Haidong Zhang, et al. 2024. Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay. *TVCG* 30, 1 (2024), 109–119.
[10] Tarique Siddiqui, Albert Kim, John Lee, et al. 2016. Effortless Data Exploration with zenvisage: An Expressive and Interactive Visual Analytics System. *PVLDB* 10, 4 (2016), 457–468.
[11] Manasi Vartak, Sajjadur Rahman, Samuel Madden, et al. 2015. SEEDB: Efficient Data-Driven Visualization Recommendations to Support Visual Analytics. *PVLDB* 8, 13 (2015), 2182–2193.
[12] Yupeng Xie, Yuyu Luo, Guoliang Li, and Nan Tang. 2024. HAIChart: Human and AI Paired Visualization System. *PVLDB* 17, 11 (2024), 3178–3191.
[13] Yupeng Xie, Zhiyang Zhang, Yifan Wu, et al. 2025. VisJudge-Bench: Aesthetics and Quality Assessment of Visualizations. *arXiv preprint arXiv:2510.22373* (2025).